Journal of Building Engineering 95 (2024) 110221

Contents lists available at ScienceDirect

Journal of Building Engineering

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

Evaluating different CFD surrogate modelling approaches for fast and accurate indoor environment simulation

Lige Zhao [a,1], Qi Zhou [a,1], Mengying Li [b], Zhe Wang [a,*]

[a] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong Special Administrative Region

[b] Department of Mechanical Engineering & Research Institute for Smart Energy, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region

ARTICLE INFO

Keywords:
Surrogate model
Proper orthogonal decomposition (POD)
Artificial neural networks (ANN)
Computational fluid dynamics (CFD)
Indoor environment quality (IEQ)
Fast prediction

ABSTRACT

Indoor Environment Quality (IEQ) holds significant importance in building design and operation, and the simulation of indoor environments playing a crucial role in enhancing IEQ. Although Computational Fluid Dynamics (CFD) has been widely employed for simulating building environments, it is computationally demanding, particularly for large spaces. To tackle this challenge, we conducted a systematic evaluation of three surrogate models for accelerating CFD: Proper Orthogonal Decomposition (POD), Artificial Neural Networks (ANN), and a combined POD-ANN approach. Our evaluation criteria focused on assessing the model accuracy, the model size, computational time and extrapolation ability. A validated CFD case model and a real campus building are employed for model evaluation. The findings demonstrate that the top five modes can reconstruct the original data matrix accurately, and the POD-ANN significantly reduces model complexity and computation time by reducing the number of parameters in the neural network, the POD-ANN parameters is only 0.14 % of the ANN, and computation time is reduced by 63 %. In addition, the combination with ANN helps increase the extrapolation ability of POD significantly. In conclusion, this research proves that the POD-ANN can enhance the efficiency of CFD calculations with the advantages of both ANN and POD. By applying the POD-ANN to predict indoor temperature, we achieve faster predictions without compromising model accuracy, and an excellent extrapolation ability is achieved. This approach also reduces model complexity, highlighting its practical value for indoor environment prediction, particularly for large and complicated spaces.

1. Introduction

1.1. Background

Individuals spend 90 % of their time indoors throughout their lives [1], making the indoor environments significantly impactful on human life. Improving Indoor environment quality (IEQ) has always been a key issue in building design and operation [2]. IEQ includes two important aspects: indoor air quality (IAQ) and thermal comfort. Research has shown that IAQ significantly impacts human health [3]. Poor indoor air quality can lead to various diseases such as airborne infectious diseases, allergies, asthma, and Sick Building

* Corresponding author.
E-mail address: cezhewang@ust.hk (Z. Wang).
[1] These two authors contribute equally to this paper.

https://doi.org/10.1016/j.jobe.2024.110221
Received 1 May 2024; Received in revised form 10 July 2024; Accepted 14 July 2024
Available online 15 July 2024

Syndrome [4]. IEQ can even affect productivity [5]. For example, changes in temperature and lighting cause potential alterations in workers' physiological and psychological performance [6], which can lead to increased medical costs and sick leave [7]. In addition to enhancing the comfort and health of building occupants, reducing building energy consumption is another important subject in studying the indoor environment. The creation of IEQ is closely related to the overall energy consumption of a building [8]. How to reduce energy consumption while ensuring high IEQ is a critical issue. Given the complexity of the building environment and the unpredictability of indoor human behavior, this problem becomes particularly complex.

Simulating the indoor environment is the first step to manage a high-quality indoor environment. An effective model can assist researchers in better understanding the state and potential changes of the built environment under various boundary conditions. Moreover, this process eliminates the need for extensive experiments and the installation of various sensors to measure IEQ parameters, thereby saving substantial human, time, equipment, and financial resources [9]. Accurate predictions can help anticipate changes in IEQ, enabling the preplanning of building equipment operation strategies, such as HVAC (Heating, Ventilation and Air Conditioning) systems in buildings. This facilitates better control over IEQ, providing a comfortable and healthy environment while reducing building operational energy usage.

Among the simulation tools for IEQ, Computational Fluid Dynamics (CFD) is a powerful and widely used method. It has been utilized for decades in the indoor environment modeling and applied in various complex ventilation systems or buildings [10]. CFD can accurately simulates airflow [11], heat transfer [12], and particle diffusion [13]. Gilani et al. [14] conducted systematic study on the temperature stratification of room displacement ventilation using 3-dimensional steady Reynolds Averaged Navier-Stokes (RANS) CFD simulations. T. van Hooff et al. [15] used the same method for an in-depth investigation of CO_2 concentration decay in a large semi-enclosed stadium after a concert. Beyond this, a wealth of valuable literature has been accumulated on research into IAQ [16–19] and natural ventilation [20–23] using CFD. From its inception to its widespread application today, the indispensability of the CFD in the IEQ field is evident, as also emphasized in literature reviews [24–26]. As summarized earlier, CFD is a high-fidelity modeling approach grounded in detailed physical principles, capable of successfully executing detailed predictions of various airflow and temperature distributions [27]. CFD is based on numerical analysis to solve fluid flow problems, involving solving partial differential equations (PDE) related to mass, momentum, energy, and turbulence [24]. CFD modeling typically involves the construction of a 3D model, mesh generation and computation. The solving process is based on the principles of momentum conservation, mass conservation, and energy conservation to solve for the temporal and spatial variations of the fluid within each mesh cell. But this process consumes significant computational resources. With the optimization of software computation methodologies and the enhancement of hardware performance, the computational costs of CFD have become almost negligible compared to several decades ago, however, this computational load is still too large for control-oriented predictions. Real-time simulation is required in the control of many building systems, posing a substantial challenge for current CFD methods.

1.2. Existing work

For high computational demands and lengthy computation processes, surrogate models present an ideal solution. To address the issue of time-consuming CFD, various studies have proposed different surrogate model strategies, including data-driven approaches and reduced-order models [28]. One of the most representative methods in data-driven approaches is Artificial Neural Networks (ANN) [29], while the most prominent among reduced-order models is the Proper Orthogonal Decomposition (POD) [30].

1.2.1. ANN to accelerate CFD

ANN is one of the most representative supervised learning algorithms in deep learning [29], endowed with powerful fitting capabilities that can approximate arbitrarily complex functional relationships. It also exhibits strong fault tolerance and robustness, as well as excellent extrapolation ability [31]. With the advancement of computer hardware, ANNs have become increasingly well-suited for handling large datasets and complex prediction tasks. At present, ANNs have been extensively applied in areas such as building energy consumption prediction [32], and optimization of HVAC systems [33]. In recent years, a growing number of researchers have focused on the application of ANN in the field of CFD. Calzolari et al. [34] conducted a literature review on the use of ANN in CFD simulations related to the built environment. They highlighted that the low computational cost of ANN can serve as a substitute for or an enhancement to the expensive CFD processes, enabling rapid predictions of indoor airflow. Moreover, CFD generates large volumes of data, providing an excellent data foundation for training ANNs. For instance, Zhou et al. [35] proposed a CFD surrogate model based on neural networks to predict indoor temperature and air flow speed. Under this method, the original CFD calculation process which took 8 h was compressed to 1.7 h, reducing the time by 80 %. In their other research [36,37], they compared the performance of different deep neural network models in indoor airflow prediction, and investigate other issues such as data preprocessing and weight-initialization methods of ANN-CFD surrogate models. However, purely data-driven methods like ANNs still have some issues. For instance, they are unable to understand the inherent physical knowledge in the learning data, and the need for extensive data training to achieve good model accuracy.

1.2.2. POD to accelerate CFD

POD, also known as Principal Component Analysis (PCA), is an effective modal analysis method that has been widely applied in fluid dynamics, mechanics, signal processing, control systems and deep learning. POD can extract meaningful patterns from high-dimensional fluid dynamics [38,39], thus implementing model reduction [40,41]. The application of the POD in flow field modal decomposition was first proposed by Lumley in 1970 [42]. After decades of development, it has been widely applied to various flow field problems [38]. Taira et al. [43] proposed a new immersed boundary method based on POD, whose structure is algebraically the same as the traditional fractional step method, used for calculating incompressible flow with prescribed surface motion. Sirovich et al.

[44] studied turbulence issues in flow fields based on the snapshot-POD. Furthermore, various complex flow field problems [39] such as cylinder wakes, wall-bounded flows, airfoil wakes, cavity flows, etc., have all been studied using POD. In recent years, scholars have extended the application of this method to other fields, such as building environments predictions. Li and colleagues [45–48] have applied the POD to the fast simulation of indoor thermal environments and CO_2 concentration levels, integrating this approach with indoor environmental control algorithm strategies. The control method based on the POD surrogate model is characterized by its low computational cost and high spatial resolution. Building upon these studies, Luo et al. [49] have expanded the application of POD under more complex boundary conditions, enriching the research experience related to the prediction of indoor non-uniform environments. It is important to note that the prediction tasks based on POD necessitates the use of additional methods to complete the prediction process. The most commonly used technique is interpolation [50] such as Radial Basis Function (RBF) [49].

1.2.3. Combination of POD and data driven methods

Interpolation-based POD can achieve good predictive performance within a specific data range. However, this method has poor extrapolation capability because interpolation techniques are unable to predict data outside the known range. Compared to interpolation methods, ANNs, with their outstanding generalization capability, not only perform well on known data but also achieve accurate predictions for data outside the training set. Qiu and colleagues [51] have utilized a Kriging surrogate model to form a mapping between boundary conditions and POD bases, which has demonstrated superior performance in both subsonic and transonic flow field predictions, requiring less than 1/200 of the time taken by traditional CFD solvers. Jia et al. [52] employed an alternative Backpropagation Neural Network (BPNN) algorithm to conduct similar research, with findings indicating that BPNN can also achieve enhanced computational efficiency while maintaining high model accuracy. In the research conducted by Huang et al. [53], POD was also employed in conjunction with BPNN to predict structural wind loads. The method demonstrated precision and robustness that were significantly superior to those of the inverse square distance weighting (ISDW) method.

1.3. Scope and objectives

In previous studies, both POD and ANN have been extensively explored in the realm of predictive modeling for indoor environment applications. However, our literature review identified two research gaps. First, the research on integrated POD with ANN is still limited. Second, the systematic evaluation of different surrogate modeling approaches has not been studied. To address these research gaps, this study, originating from a room case that has been experimentally validated, evaluates the application of the POD, ANN, and the combination of POD and ANN for the rapid prediction of non-uniform indoor thermal environments with mechanical ventilation and internal heat sources. The comparison focuses on the model accuracy, computational efficiency, and generalizability. The remaining of this paper is organized as follows: we will first introduce the methodology we proposed in Section 2, including the workflow adopted in this paper. And the detail of the test case will be given in Section 3. Next, we will discuss the results in Section 4, including the results of three methods and the comparison of them. At last, we will present the conclusions in Section 5.

2. Methodology

This section introduced the three CFD surrogate models adopted in this research, including *direct POD*, *direct ANN* and combined *POD-ANN*. The methodological outline and workflow are referred to in Fig. 1. Notably, there are two ANN architectures adopted in this research, including direct ANN and POD-ANN, and the differences between them can be found in the figures the relevant sections. As depicted in Fig. 1, the three models employ the same dataset, which has been consistently partitioned into training, validation, and test

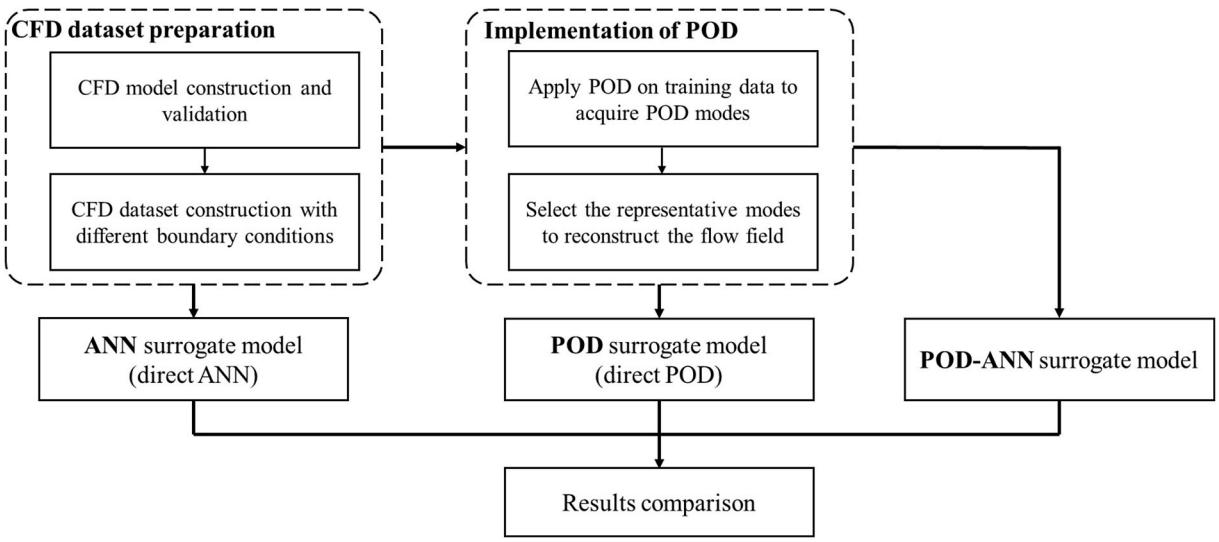


Fig. 1. Methodology outline.

sets. Each data in the dataset represents the simulated outcomes under varying boundary conditions within the CFD model.

2.1. POD surrogate model

2.1.1. Theory

The mathematical principle underlying POD is the orthogonal decomposition of a matrix, which allows the original matrix to be decomposed into eigenvalues and eigenvectors. In the modal decomposition of a flow field, each eigenvector can be regarded as a characteristic of the original flow field, while the eigenvalues represent the intensity of these characteristics. The eigenvalues typically decrease very rapidly, indicating that only the first few eigenvectors are usually sufficient to reconstruct the original flow field. There are typically three methodologies for applying POD [38]: Classical POD (also referred as Spatial POD), Snapshots POD, and Singular Value Decomposition (SVD). All three methods share a common principle: they decompose data matrices and extract the primary modes to reduce the dimensionality of the original data. Compared to the Classical POD, the Snapshots method can further reduce the necessary computational and memory resources, particularly well-suited for high-dimensional fluid flow, while the SVD based POD exhibits superior robustness when dealing with roundoff errors. In this study, the adopted method is SVD. As illustrated in Fig. 2(a), any matrix $A \in \mathbb{R}^{m\times n}$ can be decomposed into the three matrices as

$$A = USV^T \tag{1}$$

where $U \in \mathbb{R}^{m\times m}$ is an orthogonal matrix whose column vectors are the left singular vectors and form an orthogonal basis for the column space of the original matrix. $S \in \mathbb{R}^{m\times n}$ is a diagonal matrix whose elements are singular values arranged in descending order, representing the square roots of the non-zero eigenvalues of A^TA or AA^T. $V \in \mathbb{R}^{n\times n}$ is an orthogonal matrix whose column vectors are the right singular vectors and form an orthogonal basis for the row space of the original matrix. As depicted in Fig. 2(b), assuming that $A = A_nV^T$, where $A_n = US$, and A_n can be interpreted as stretching or scaling of the original matrix in the direction of the column space, while V^T is the rotation or reflection in the direction of the row space. With these conclusions, V^T can be considered a mode classifier, while A_n can be seen as a feature energy of the original data matrix. Then the top modes are selected for reconstructing the original matrix, as shown in Fig. 2(c).

2.1.2. POD data training: decomposition

The training of POD is illustrated in Fig. 3. The training dataset is denoted by a matrix U, where n represents the number of boundary conditions. The variable m denotes the size of the data, namely the mesh number in CFD model. Any row of U represents the temperature distribution data under each boundary condition, obtained by flattening the temperature data of all the mesh. Applying SVD to matrix U results in the amplitude matrix A_n and mode matrix V^T. In this study, the selection of modes number m' depends on the feature retention capability of the leading modes for the training set matrix. The selection entails starting with the first-order mode and comparing the results of the reconstructed data with the original. The number of modes is gradually increased until the mean absolute error (MAE) between the original and reconstructed temperature distribution is small enough.

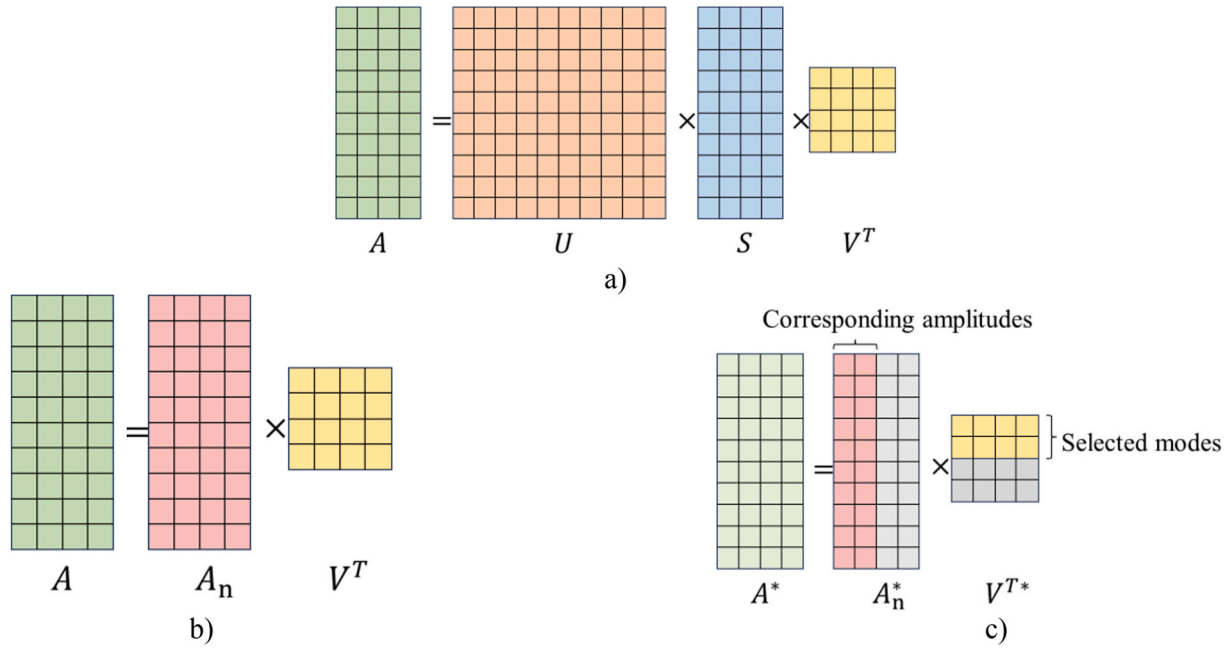


Fig. 2. Theory of POD with SVD: a) Any given data matrix A can be decomposed by SVD, b) V^T is the mode matrix, and U × S is the amplitude matrix, c) Select the top modes to represent the original data matrix A for the purpose of dimension reduction.

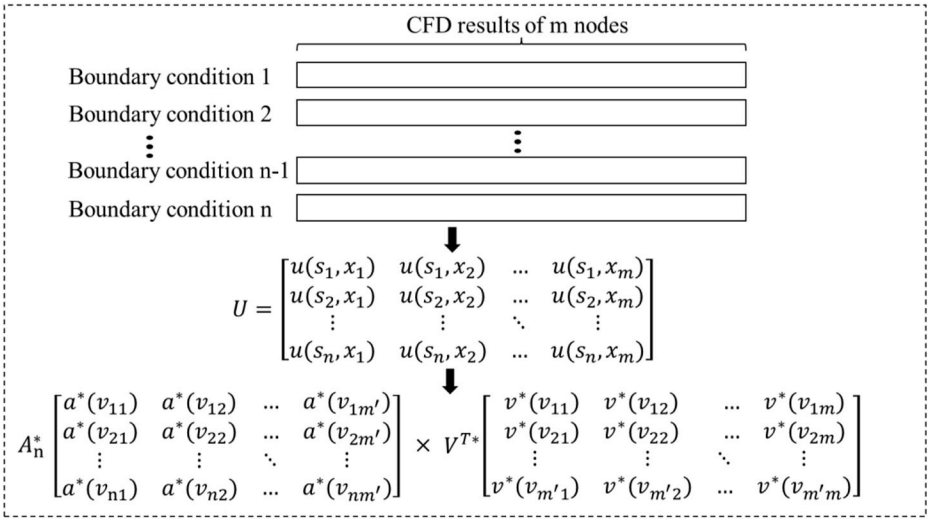


Fig. 3. Decompose the training data matrix and select top m' modes, where $m' < \min(m,n)$.

2.1.3. POD prediction: interpolation

After the decomposition of the training set, the interpolation is implemented on A_n^*, and the amplitude matrix $A_n^{*'}$ of the test set is computed according to the interpolation. Then $A_n^{*'}$ is multiplied by V^{T*} for prediction. RBF served as the interpolation method in this study. RBF is a scattered data interpolation technique that offers a smooth approximation based on the concept of distance for data points in multidimensional space. In RBF, for each point within the interpolation region, its function value can be approximated by a linear combination of radial basis functions of the surrounding known data points. For a set of scattered data $(x_i,y_i), i = 1,2,\dots,N$, the goal of RBF is to find a function $f(x)$ such that $f(x_i) = y_i$ for all data points, and $f(x)$ is smooth throughout its entire domain. The form of $f(x)$ is typically expressed as

$$f(x) = \sum_{i=1}^{N} \lambda_i \Phi(||x - x_i||) \tag{2}$$

where Φ is the selected radial basis function, $||x - x_i||$ is the Euclidean distance of x and x_i, λ_i is the undetermined coefficient. The radial basis function $\Phi(r)$ is a function of distance r that satisfies radial symmetry, meaning that the value of $\Phi(r)$ depends solely on the magnitude of r and is independent of direction. There are several commonly used radial basis functions, and the function adopted in this study is multiquadric function.

$$\Phi(r) = \sqrt{r^2 + c^2} \tag{3}$$

where r the distance and c is a constant that can be adjusted according to the data. The determination of λ_i typically achieved by solving a system of linear equations formed by the interpolation conditions $f(x_i) = y_i$, which can be expressed as follows.

$$\begin{bmatrix} \Phi(||x_1 - x_1||) & \cdots & \Phi(||x_1 - x_N||) \\ \vdots & \ddots & \vdots \\ \Phi(||x_N - x_1||) & \cdots & \Phi(||x_N - x_N||) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \tag{4}$$

which can be solved sing standard linear algebra techniques such as LU decomposition, Gaussian elimination, or singular value decomposition (SVD). Once λ coefficients are determined, a new amplitude matrix y can be calculated for each set of new boundary conditions x through the function $f(x)$.

2.2. ANN surrogate model

A typical feed-forward neural network consists of an input layer, multiple hidden layers, and an output layer, with neurons in each layer. These neurons connect the previous layer with the next one, forming a structure like the connections of neurons in the human brain. The connection method between each layer's neurons involves the signal of a single neuron receiving the weighted sum of the data from each neuron in the previous layer and a bias term (optional), then feed the value into an activation function to provide the neuron output to the next layer. This can be represented as

$$y = f\left(\sum_{i=1}^{n} W_i x_i + b\right) \tag{5}$$

where W_i is the weight of the i-th neuron in the previous layer, b is the bias term, and $f(x)$ is the activation function. Several activation

functions are available, but nonlinear ones are typically used in complex prediction scenarios. Once the data has been transmitted to the output layer, one propagation is complete. To enhance the performance of the neural network, the error between the results and ground truth is calculated and then backpropagated to the input layer. The weights and bias from the previous propagation are adjusted and updated using a gradient descent algorithm. The hyperparameters of neural networks include the number of layers, the number of neurons in each layer, the activation function, and the gradient descent algorithm. The structure of the direct ANN is shown in Fig. 4. This neural network is trained directly through boundary conditions to predict temperature results. The inputs are the supply air temperature T_{in}, wall temperature T_w, and heat source temperature T_h. Notably, the variables T_{in}, T_w, and T_h are preprocessed before inputting into the neural network to mitigate the impact of the feature scale and enhance the model robustness. In this study, standardization is adopted as the preprocessing method and can be expressed as

$$T^* = \frac{T - \mu_T}{\sigma_T} \tag{6}$$

where μ_T, σ_T represent the mean value and the standard deviations of the temperature data. The predicted temperature distribution is also preprocessed with the same process.

In neural network training, hyperparameters notably affect the performance of the neural network and usually require experimentation and search to determine. In this study, Optuna, an open source hyperparameter optimization framework, is used for optimization. The hyperparameters and their search range are shown in Table 1.

2.3. POD-ANN surrogate model

The workflow of POD-ANN is depicted in Fig. 5. Specifically, it involves establishing a neural network with an architecture similar to Fig. 4. In this neural network, the modal intensities matrix corresponding to the leading modes serves as the training target. Hence, the dimension of the output layer is m' according to the selected number of modes. Like the direct POD approach, the training of POD-ANN also encompasses the decomposition, and the neural network is employed to establish a mapping between the amplitude matrix V^{T*} of the training set and the boundary conditions. Then use the trained neural network to predict the amplitude matrix of the new boundary conditions. Finally, multiply the results with the V^{T*} to obtain the predicted temperature distribution.

2.4. Model summary

The inputs of the three surrogate models are the same, but the training processes and model outputs are different. Based on the predictive objectives of this study and the principles of these models, the key information for each model is organized as shown in Table 2. The inputs for models are combinations of boundary conditions, namely supply air temperature, heat source temperature, and wall temperature values. For direct ANN, the model forges a direct mapping between the boundary conditions and the temperature distribution, and the output is a one-dimensional array containing the temperatures in the room, each data point corresponding to a three-dimensional coordinate. For both the direct POD and POD-ANN models, the modal decomposition is required, then the amplitude matrix from the decomposition becomes the prediction target. The output is the amplitude matrix corresponding to each

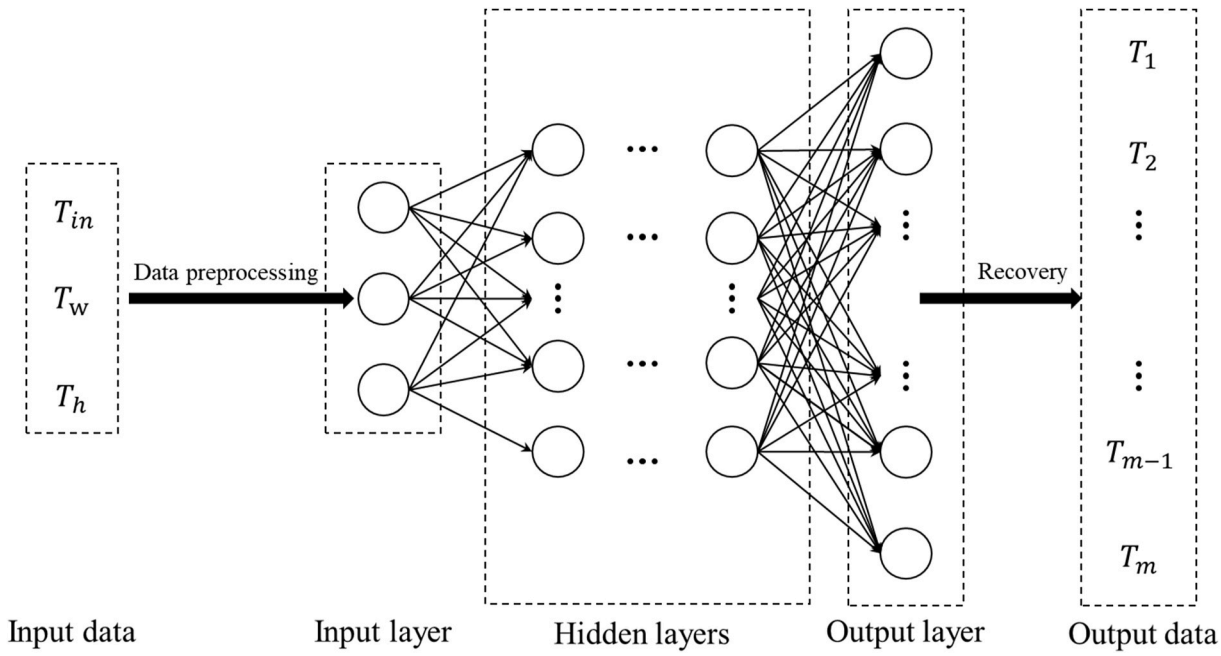


Fig. 4. Neural network architecture of direct ANN.

Table 1
Hyperparameters optimization with Optuna.

Hyperparameters		
Fixed	Activation function	ReLU
	Gradient descent algorithm	Adam
	Batch size	1
	Training epochs	2000
	Number of neurons of input layer	3
	Number of neurons of output layer	4096
To be optimized	Number of hidden layers	1–5
	Number of neurons of each hidden layer	4–64
	Dropout rate	0.1–0.5
	Learning rate	0.00001–0.1

boundary condition. To obtain the predicted temperature, it is necessary to multiply the amplitude matrix by the selected modes. As can be inferred in Section 2.3, under the POD-ANN framework, the role of the neural network is to substitute the interpolation. Accurate predictive outcomes can be achieved by harnessing the neural network's potent regression capabilities. The construction, training, and prediction of the ANN model are all completed on a Windows desktop with an 11th generation Intel i5-11400F @ 2.60 GHz CPU, which possesses a configuration of 6 cores and 12 threads, while the system is equipped with 16 GB of RAM. The programming language used is Python, version 3.9, and the model is implemented using Keras, a widely used deep learning library.

2.5. Performance evaluation criteria

2.5.1. Model accuracy

The evaluation of model accuracy includes the comparison of temperature heatmap and error statistical indicators include the Mean Absolute Error (MAE) and the Normalized Root Mean Squared Error (NRMSE), which are calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^{m}\left(T_{pred} - T_{CFD}\right)^2}{\sum_{i=1}^{m}\left(\overline{T_{CFD}} - T_{CFD}\right)^2} \tag{7}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|T_{pred} - T_{CFD}\right| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(T_{pred} - T_{CFD}\right)^2} \tag{9}$$

$$NRMSE = \frac{\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(T_{pred} - T_{CFD}\right)^2}}{T_{CFD_{max}} - T_{CFD_{min}}} \tag{10}$$

where T_{pred} is the predicted value, T_{CFD} and $\overline{T_{CFD}}$ are the CFD results and the average, $T_{CFD_{max}}$ is the maximum value of CFD results, while $T_{CFD_{min}}$ is the minimum.

2.5.2. ANN architecture complexity

Model complexity is critical for neural network models. Given equivalent predictive accuracy, models with lower complexity conserve more memory and hardware resources, thereby enhancing computational efficiency. In this research, the architectures and model hyperparameters of the two neural networks in direct ANN and POD-ANN will be compared.

2.5.3. ANN computation time

Discussions of computation time within this study are confined to the duration required for making predictions and exclude the time spent on data retrieval and the training process. This is because the training phase consumes substantially more time than prediction, and the optimization process with Optuna is stochastic, leading to significant variability in outcomes with each execution.

2.5.4. Extrapolation ability

The extrapolation capability of a model represents its prediction performance on an unknown dataset and is one of the key criteria for assessing a model's prediction capacity. In this study, the extrapolation abilities of the three surrogate models are verified by partitioning the data into distinct training and testing sets.

3. Test case

3.1. CFD modeling and validation

To ensure the reproducibility of this work, we use an open-source validated CFD model [54] as our case. The geometric structure of

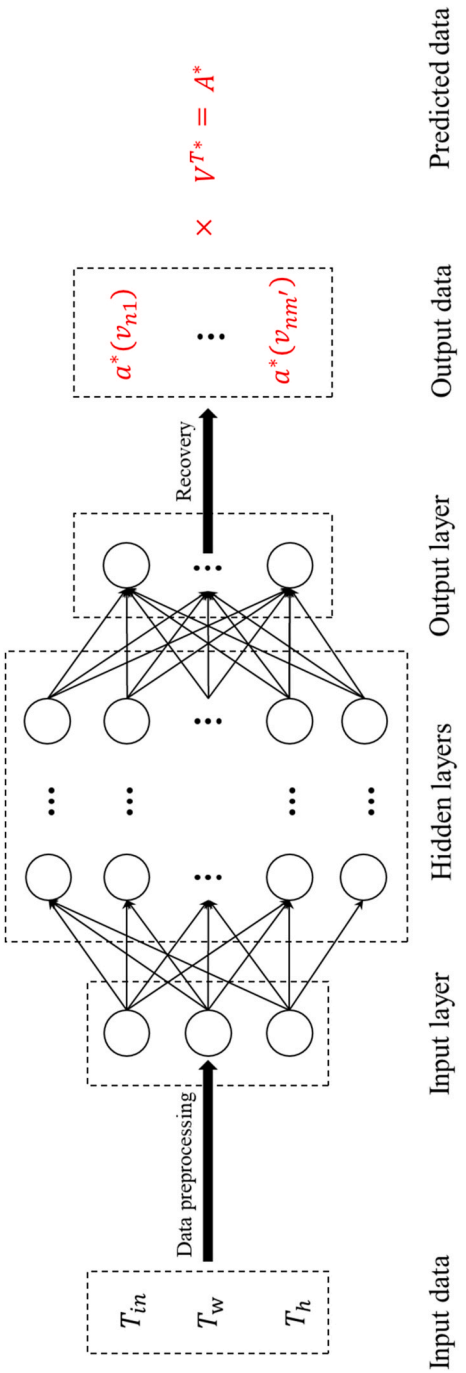
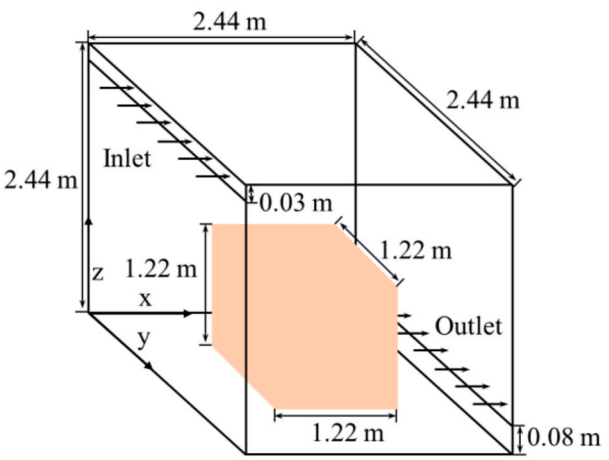


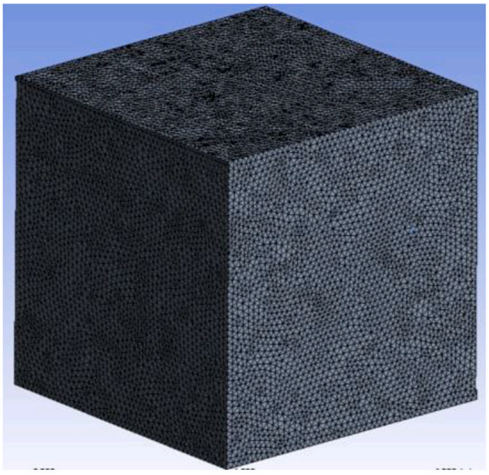
Fig. 5. Neural network architecture and workflow of POD-ANN.

Table 2
Comparison of three surrogate models.

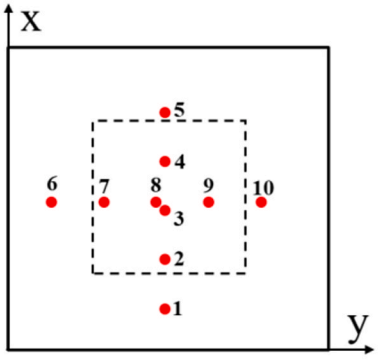
	Direct ANN	Direct POD	POD-ANN
Inputs	Temperature of supply air, heat source and walls	Temperature of supply air, heat source and walls	Temperature of supply air, heat source and walls
Outputs	Temperature data	Amplitude matrix	Amplitude matrix
Prediction method	Neural network regression	Interpolation	Neural network regression
Need for reconstruction	No	Yes	Yes

the room model for this case, as shown in Fig. 6(a)–is a cubic laboratory equipped with an internal heat source and mechanical ventilation. The room size is 2.44 m (×) 2.44 m (×) 2.44 m and built with insulation materials. In the model room, a 1.22 m (×) 1.22 m (×) 1.22 m cubic box represents furniture, with a heat source inside the box generating complex indoor airflow through heat dissipation to the air. The experiment assumes that all seats and occupants are concentrated within this box, using 16 light bulbs to produce a 700W heat source, equivalent to the sensible heat generated by 10.5 occupants. Four fans inside the box ensure uniform heating of the box's surface. Under these conditions, the surface temperature of the box can be stabilized at 36.7 °C, which is comparable to human body temperature. The experiment includes three test cases to account for the complexity of flow field characteristics. The first case is isothermal forced convection without a heat source, the second case involves the box without a heat source, and the third case places a 700W heat source inside the box. This study selects the most complex third test case to investigate the performance of different surrogate models under complex flow field conditions. In the experiment, the ceiling temperature is 25.8 °C, the floor temperature 26.9 °C, and the temperature for the other walls is 27.4 °C. An air inlet, set at a height of 0.03 m and equal to the width of the wall, is installed close to the ceiling on one side to supply indoor ventilation. An outlet, with a height of 0.08 m and again equal to the wall's width, is set close to the floor on the opposite wall to facilitate whole-room air circulation. The inlet temperature is 22.2 °C with a wind speed of 1.36 m/s. They established a total of 512 measurement points in the laboratory to compare experimental





(a)



Point	X_Coordinate	Y_Coordinate
1	0.2286	1.2192
2	0.6858	1.2192
3	1.1430	1.2192
4	1.6002	1.2192
5	2.0574	1.2192
6	1.2192	0.2286
7	1.2192	0.6858
8	1.2192	1.1430
9	1.2192	1.6002
10	1.2192	2.0574

(b)

Fig. 6. Brief introduction of the test case: (a) geometry and CFD meshing of the model room, (b) data points coordinate.

and simulation results. In their work, data points labeled as Point 1 to Point 10 are mentioned, as shown in Fig. 6(b), and they make public the test results for Point 1, 3, 5, and 6.

Parameters of the CFD model are indicated in Table 3, the air is treated as incompressible ideal gas. The walls and heat source surfaces are specified as Dirichlet boundary conditions. The inlet is defined by velocity boundary conditions, allowing for the adjustment of supply air velocity and temperature. The outlet is set as a pressure boundary condition with the exit pressure set to 0. The RNG k−ε turbulent model is applied, and the Boussinesq approximation is utilized to address buoyancy effects. A second-order upwind scheme is employed for solving the control equations. The mesh size is 0.05 m and a total mesh number is 77,387. This study simplified the laboratory's internal heat source in the CFD model by replacing the original 700W internal heat source with a cubic box that maintains a constant surface temperature of 36.7 °C. To validate the model, simulations are conducted using parameters consistent with the experiment in Ref. [54].

CFD simulations are executed using the widely adopted commercial software FLUENT, version 2022 b1, running on the same desktop mentioned in Section 2.4. The temperatures of the CFD result are presented along the Z-axis, from top to bottom, under different x and y coordinates at each data point. The validation is illustrated in Fig. 7. The square points in the image represent the measured temperature, while the dashed lines denote the CFD results. Both temperature data and Z-axis height in the figure are normalized. The vertical coordinate represents the ratio of the height to the total wall height (Z/L), while the horizontal coordinate is shown as

$$T^{*} = \frac{T - T_{low}}{T_{high} - T_{low}} \tag{11}$$

where T, T_{low}, T_{high} represent the test temperature, the lowest temperature 22.2 °C and the highest temperature 36.7 °C. The figure shows that the deviation between the CFD simulation results and the experimental results is minimal, with the simulation closely matching the test data at most data points. The error metrics comparing CFD results with experimental tests are presented in Table 4. The table shows the results for MAE, RMSE, and NRMSE. Among the four data points, the highest MAE is 0.031, the highest RMSE is 0.044, and the highest NRMSE does not exceed 14 %. The simulation results suggest that the CFD dataset used in this study can accurately represent the temperature distribution within a real building under the given geometric structure, thereby affirming the validity of this study. To alleviate computational cost, the results from the CFD simulations are dimensionality reduced. Specifically, the total number of nodes in the entire room, initially 77,387, is reduced to 4096 (16 × 16 × 16), which will not affect the accuracy of the model [35]. In subsequent research, comparisons between different surrogate models are all predicated on the results obtained from this CFD model.

3.2. Boundary conditions definition

The training and prediction of surrogate models require the preparation of a sufficient dataset in advance. The generation of dataset involves altering the boundary conditions of the CFD model each time to produce multiple sets of simulation results. The objective of this study is to predict the temperature field, without considering the distribution of velocity. Therefore, the supply air velocity is kept constant at 1.36 m/s under any boundary condition, and the heat source within the room represents the occupants heat dissipation, thus maintained at a fixed temperature of 36.7 °C. The adjustable boundary conditions are the supply air temperature and the wall temperature. The range should be large enough to ensure that each simulation result is as distinct as possible. In this study, the supply air temperature is assumed to range from 15 °C to 25 °C, while the wall temperature ranges from 10 °C to 30 °C. The sampling method employed is systematic sampling, which involves selecting samples at fixed intervals. Although different sampling methods can affect the training outcomes of the model, this issue is not discussed in this study. The sampling intervals for the supply air temperature and wall temperature are both set at 2.5 °C. Within the given ranges and sampling intervals, there are 45 boundary conditions in total.

The database is divided into training, validation, and test set. The training set is for establishing the mapping between the input and output, the validation set is to validate the training results and adjust the model hyperparameter, which is necessary for reducing overfitting and underfitting, finally, the fitting ability of the model will be evaluated on the test set. Notably, direct POD model does not

Table 3
CFD model summary.

Contents		
Geometry	Room size	2.44 m (×) 2.44 m (×) 2.44 m
	Ventilation opening	Inlet: 2.44 m (×) 0.03 m Outlet: 2.44 m (×) 0.08 m
	Heat source	1.22 m (×) 1.22 m (×) 1.22 m
Mesh	Size	0.05 m
	Number	77387
Boundary conditions	Wall temperature	Ceiling: 298.8 K Floor: 299.9 K Other walls: 300.4 K
	Heat source temperature	309.7 K
	Supply air	Temperature: 295.2 K Velocity: 1.36 m/s
Turbulence model	/	RNG k−ε turbulent model
Buoyancy effect	/	Boussinesq approximation

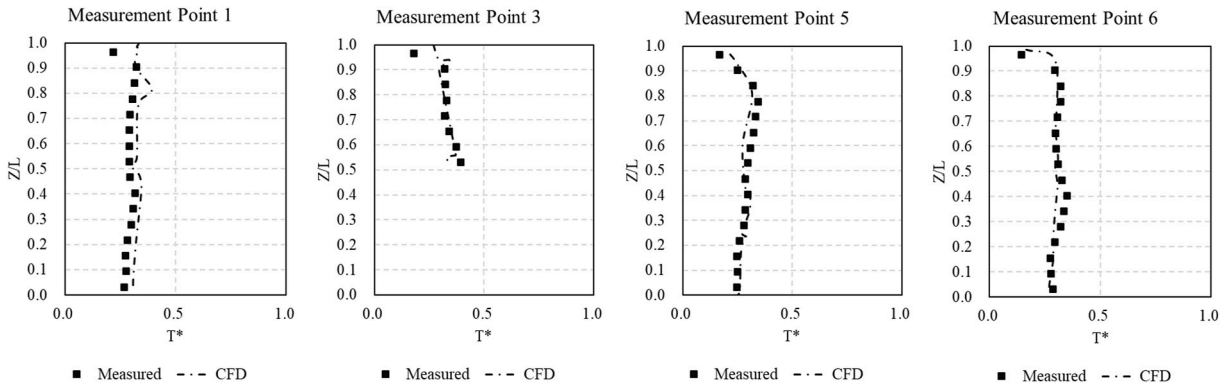


Fig. 7. CFD results validation.

Table 4
Error indicator of validation.

	Point 1	Point 3	Point 5	Point 6
MAE	0.031	0.029	0.019	0.017
RMSE	0.041	0.044	0.024	0.021
NRMSE	14.0 %	14.0 %	8.6 %	7.1 %

require validation set, as there is no hyperparameter adjustment for the interpolation, both training data and validation data are used for decomposition and modal selection. As shown in Fig. 8, 20 sets of data are designated as the training set, which includes all cases with wall temperatures of 10 °C, 15 °C, 25 °C, and 30 °C. Five sets are used as the validation set, including cases with a wall temperature of 20 °C. The remaining 20 sets are used as the test set, encompassing cases with wall temperatures of 12.5 °C, 17.5 °C, 22.5 °C, and 27.5 °C.

4. Results and discussions

4.1. Optimal modal selection for POD

The MAE between reconstructed temperature and CFD results under different number of modes is shown Fig. 9. As the number of modes m' gradually increases from one, MAE decreases sharply at first and then slows down. When the modes number is greater than 4, the MAE value is already less than 0.2 °C and reduction is very small after that, indicating that the temperature distribution of POD results is very close to the CFD simulations. In this research, the modes selected for the reconstruction is 5.2D heatmaps of temperature distribution on the XY-plane at different heights are shown in Fig. 10. Since this study is targeted towards simulation and analysis of building thermal environments, the selection of representative z-axis heights adheres to the ASHRAE 55 [55], which states that when studying indoor thermal comfort for occupants in buildings, attention should be paid to the heights of 0.1 m, 0.6 m, 1.1 m, and 1.7 m, corresponding to the height of ankle, knee, and head for seated and standing person. As the height of heat source is 1.22 m, the temperature distributions of 0.1 m, 0.6 m and 1.1 m are similar, only the heights of 0.6 m and 1.7 m are displayed in Fig. 10. An example of $T_{in} = 20°C$ and $T_w = 30°C$ is adopted in the heatmap to show the temperature distribution at different heights. Fig. 10 confirms the conclusion that when the number of modes is 5, the reconstructed temperature distribution is almost identical to the

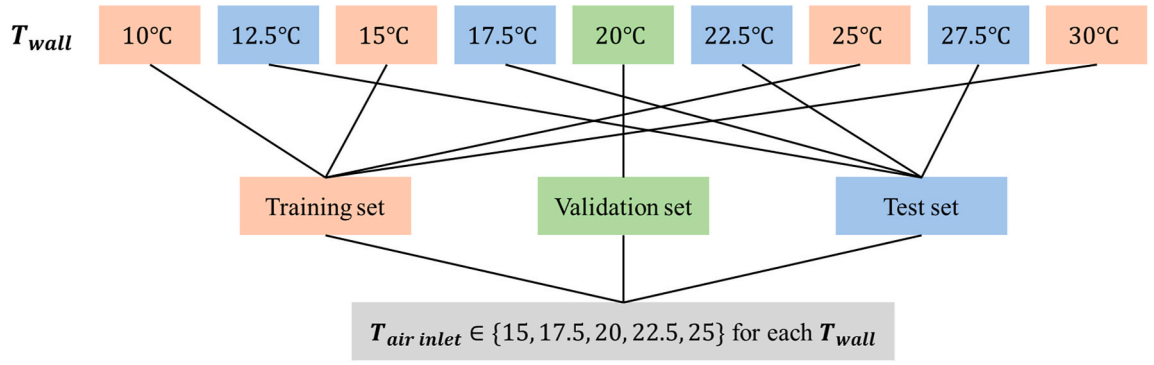


Fig. 8. Dataset partitioning: train, validation and test.

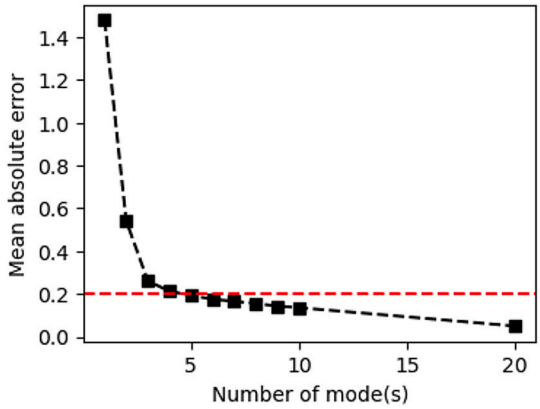


Fig. 9. Decrease of MAE with the increase of mode number.

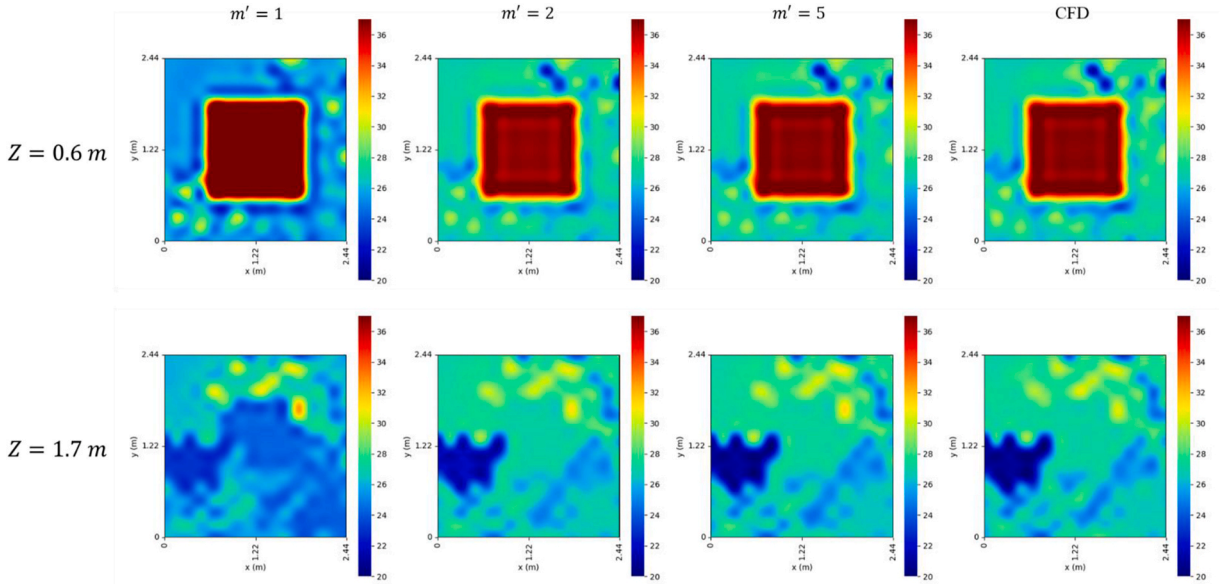


Fig. 10. Comparison of leading modes with original temperature field.

original image.

4.2. Model performance

4.2.1. Model accuracy

The predictive accuracy of the three surrogate models for the test set is summarized in Table 5, where both MAE and NRMSE represent the average values across 20 test sets. Because the outcome of each run with the neural network may vary, the results for the direct ANN and POD-ANN are averaged over 10 executions. The metrics presented in Table 5 indicate that all three surrogate models exhibit a high level of predictive accuracy, with the average temperature discrepancy between the predictions and the CFD results being within 0.4 °C, and the NRMSE within 2.5 %. This demonstrates that, under the current standards for dividing the training, validation, and test sets, all three methods possess a robust capability for predicting the temperature field. Fig. 11 showcases the two-

Table 5
Model accuracy of three methods.

Criterion	Direct POD	Direct ANN (10 runs average)	POD-ANN (10 runs average)
R-square	0.99	0.99	0.98
MAE	0.33	0.34	0.36
RMSE	0.42	0.44	0.44
NRMSE	2.3 %	2.5 %	2.5 %

dimensional temperature distribution under the conditions of $T_{in} = 25°C$ and $T_w = 17.5°C$. The heatmaps illustrate that there is virtually no difference between the predicted outcomes of the three models and the results of the CFD simulations.

4.2.2. Neural network complexity and computation time

Table 6 shows the neural network architecture and average calculation time for both the POD-ANN and direct ANN. It can be observed from the table that the total parameters of the POD-ANN have significantly decreased. Compared to direct ANN, the order of magnitude of total parameters has dropped from hundreds of thousands to hundreds. Given that the case in this study is relatively simple, there is not a significant difference in the number of hidden layers and the number of neurons per hidden layer between the two methods. The reason for this significant decrease is that the number of neurons in the output layer of the neural network is not on the same order of magnitude: the output layer dimension in the direct ANN is 4096, while it is only 5 in the POD-ANN. As described in Section 2.2, the premise for a neural network to produce accurate prediction results is to form a complex mapping between inputs and outputs. This mapping is determined by parameters such as weights and biases propagated between the layers of the neural network. That is to say, the POD-ANN can produce predictions of the same accuracy with fewer weights and biases, thus ensuring model performance while reducing model complexity. Due to the reduced complexity of the model, the average computation time for 10 runs using the POD-ANN is only 3.65 s, as opposed to 9.88 s for the direct ANN, representing a significant reduction of 63 %. This reduction is considerable and is expected to become even more pronounced with the increasing complexity of building models and flow fields, leading to further reductions in computation time.

4.2.3. Extrapolation ability

In regression problems, it is often necessary for the training set to encompass the entire data range to effectively achieve perfect predictions, meaning that new boundary conditions should fall within the scope covered by the training set. In the previous data partitioning, the training set encompass the full data range, ensuring that the test set selections do not exceed the bounds of the training set. Conversely, if the training set selection fails to cover the complete data range, the model generalizability would be in doubt, because both machine learning model and RBF are good at interpolation, but poor at extrapolation [56]. When the training set is selected as Fig. 12, repeating the workflow described in Section 2 yields statistical errors for three methods, as shown in Table 7. When the training data is set as when the T_w range is 10–17.5 °C, validation data is when T_w is 20 °C and test data covers the rest, direct ANN and POD-ANN models still maintain relatively high accuracy, but direct POD model gives a disastrous result, which demonstrates that two models with ANN can maintain high extrapolation capability under these conditions. Based on these observations, it can be concluded that within the POD-ANN framework, the neural network endows POD with enhanced capabilities for flow field reconstruction. This is attributed to the fact that the prediction of the amplitude matrix based on ANN exhibits greater robustness than the interpolation predictions derived from RBF. It demonstrates that the combination of POD with ANN outperforms the use of POD in isolation, thus indicating superior performance of the integrated approach.

4.3. Testing on a real building model

To further validate the predictive capabilities of these three surrogate models for large spatial building environments, this study constructed a 3D model of the same scale as the academic building at the Hong Kong University of Science and Technology, as illustrated in Fig. 13. The mesh number is 2040684. The building's atrium features a vertical glass curtain wall and a 45-degree angled glass ceiling on the south side, with the remaining walls in bricks. Internal heat sources include an electronic screen on the western

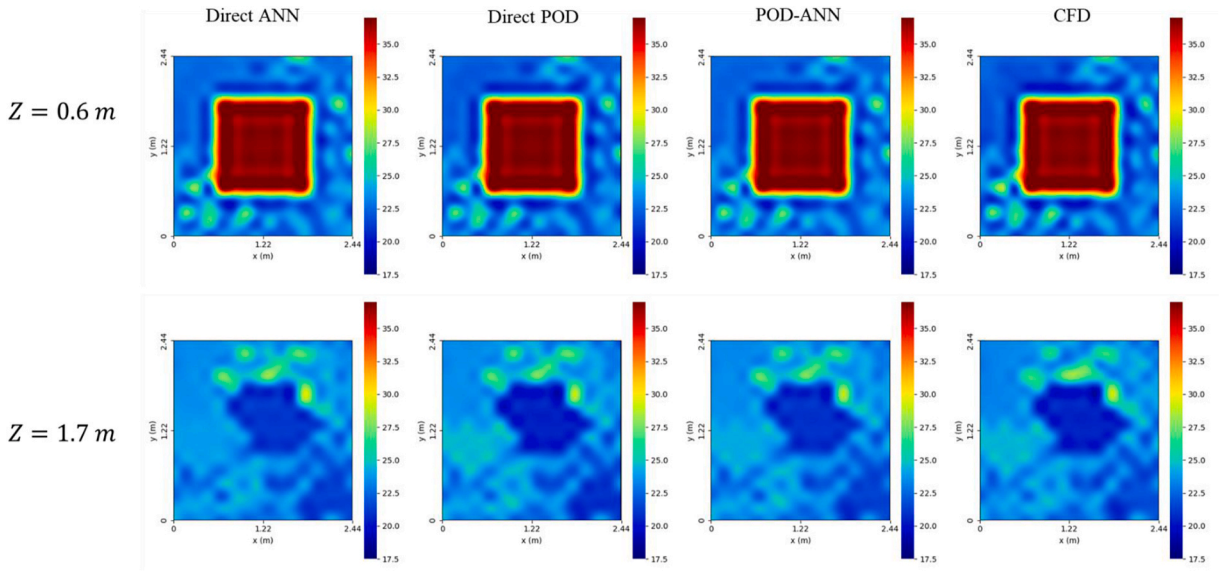


Fig. 11. Temperature distribution comparison of three methods.

Table 6
Model complexity and calculation time of direct ANN and POD-ANN.

Hyper-parameters	Direct ANN	POD-ANN
Number of hidden layers	3	1
Number of neurons of each hidden layer	38, 24, 62	40
Dropout rate	0.15, 0.26, 0.49	0.14
Learning rate	0.016	0.019
Total parameters	260282	365
Calculation time	9.88s	3.65s

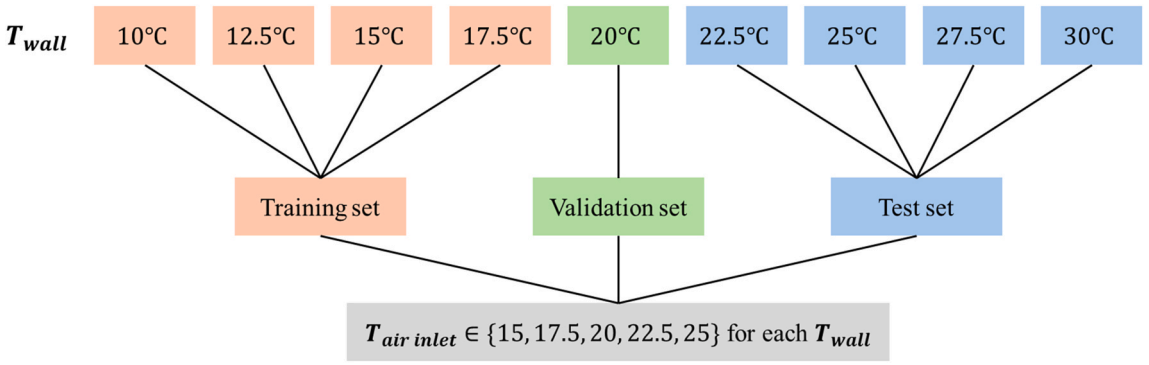


Fig. 12. Dataset partitioning: same size of training and validation set without covering the whole data range.

Table 7
Model accuracy of three models under new training set.

Criterion	Direct POD	Direct ANN (10 runs average)	POD-ANN (10 runs average)
R-square	−2.86	0.95	0.93
MAE	5.55	0.72	0.74
RMSE	6.17	1.04	1.10
NRMSE	38.4 %	5.9 %	6.2 %

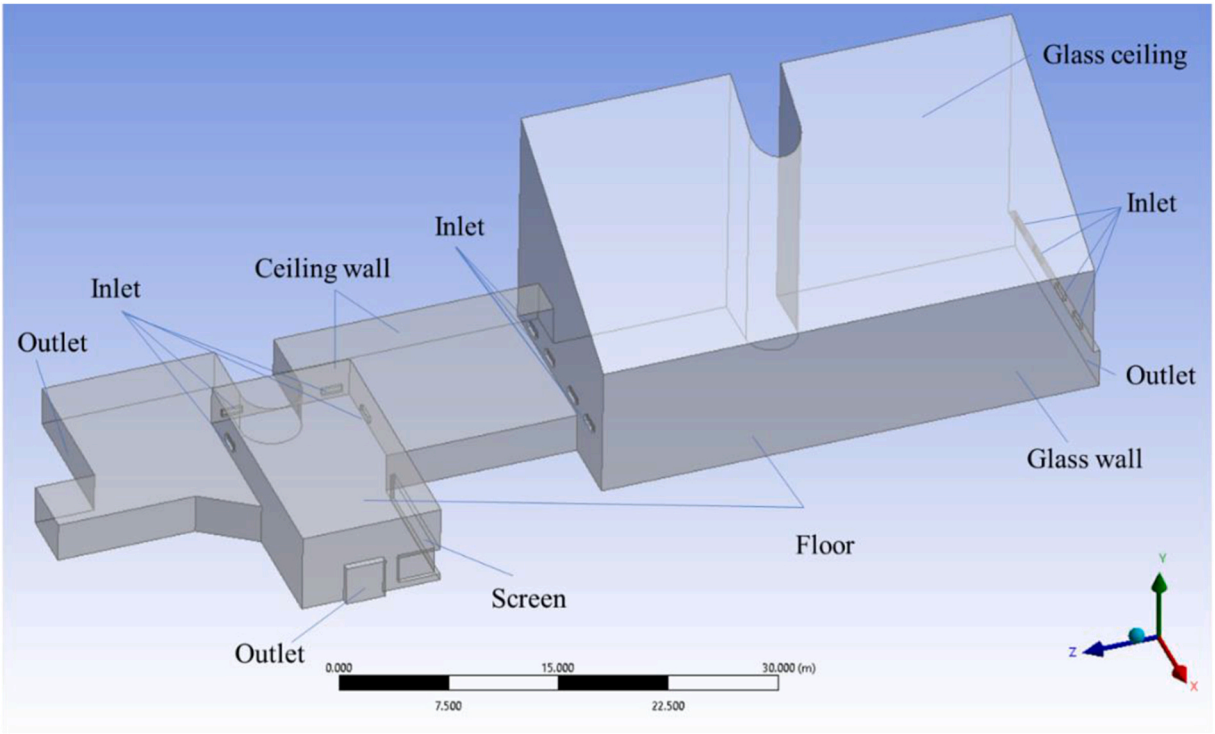
side. All air inlets in the figure have an area of 0.75 m^2, with the atrium measuring 36.25 m in length, 16.7 m in width, with the south wall at a height of 10 m, and the north wall at 20 m. The model's parameters and boundary conditions are detailed in Table 8, with the wind velocity of all air inlets being 0.8 m/s. The boundary conditions for the walls and ceiling in this model are assumed to be the same as those in Section 3, adhering to the first type of boundary conditions, i.e., fixed temperatures. In buildings with significant window-to-wall ratios, solar radiation typically exerts a substantial impact on the interior temperature. Therefore, in this model, the variables are the temperatures of the glass wall and glass ceiling, as well as the supply air temperature at the air inlets.

In accordance with the varying temperature ranges for the glass wall, glass ceiling and supply air, a total of 9 distinct combinations of boundary conditions are generated. Of these, four of them are designated for the training set, one for the validation set, and the rest for the test set, as depicted in Fig. 14.

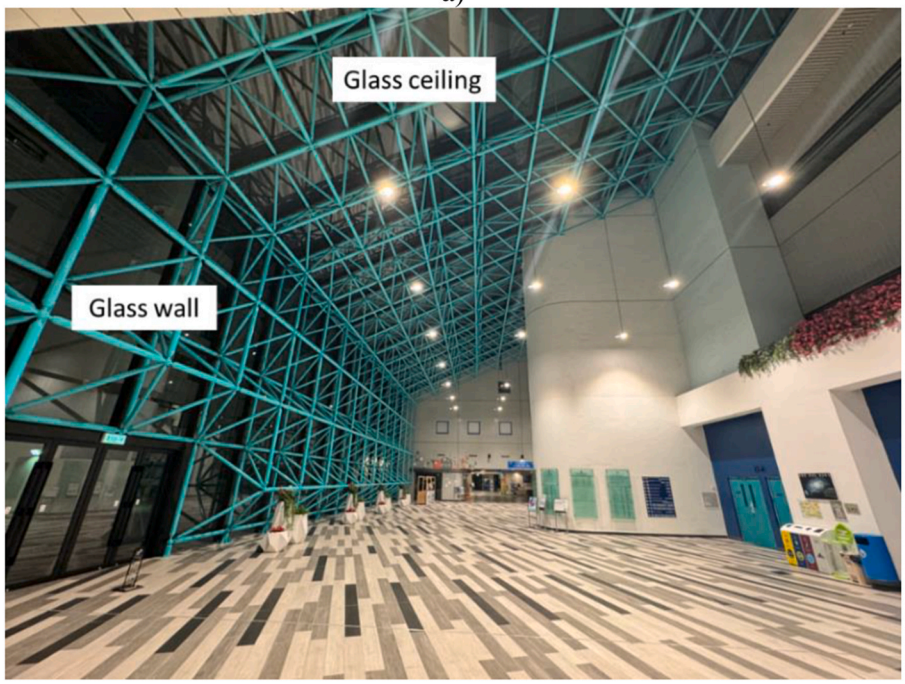
The temperature distribution of the building is predicted using the direct POD, direct ANN, and POD-ANN, respectively (see Fig. 15). The performance of these three surrogate models is presented in Table 9. From the table, it is evident that all three models exhibit a high level of prediction accuracy, with the direct ANN and POD-ANN demonstrating superior accuracy compared to the direct POD. Fig. 14 displays a comparison between the POD-ANN model predictions and the CFD results under conditions where the supply air temperature is 15 °C and the glass temperature is 50 °C. By examining the YZ cross-sections at X = 6.5 m and X = 7.5 m, it is apparent that the model is capable of accurately predicting the temperature distribution. Due to the temperature difference between the upper and lower regions, a clear stratification of temperature distribution is observed, and the surrogate models successfully reproduce this characteristic of the flow field. The models also effectively capture the distribution information of the air vent temperatures at the two positions, X = 6.5 m and X = 7.5 m, where air vents are present and absent, respectively.

5. Conclusion

The present study undertakes a comparative examination of three surrogate modeling techniques to accelerate CFD calculation under the context of simulating indoor environments. The surrogate models under scrutiny include the POD, the ANN approach, and the hybrid POD-ANN. Two distinct scenarios are considered for the comparative analysis: a validated CFD model case sourced from open literature and a real-world case involving the ground floor of a campus building. The primary performance indicators of interest encompass prediction accuracy, model complexity, computational time, and model generalizability given new boundary conditions.



a)



b)

Fig. 13. Academic building in HKUST: a) 3D model of the whole building, b) Academic corridor (atrium).

The key findings are summarized as follows:

(1) The leading modes extracted from the POD analysis of the training set accurately reproduce the flow field, indicating that only a few dominant modes are sufficient to represent the original flow field. Discarding the subsequent modes has negligible impact on information loss.

Table 8
CFD model summary of HKUST academic building.

Contents		
Invariable	Wall temperature	Ceiling wall: 303 K Floor: 293 K Screen: 303 K
Variable	Glass wall & glass ceiling temperature	303 K–323 K, at intervals of 10 K
	Supply air temperature	283 K–293 K, at intervals of 5 K

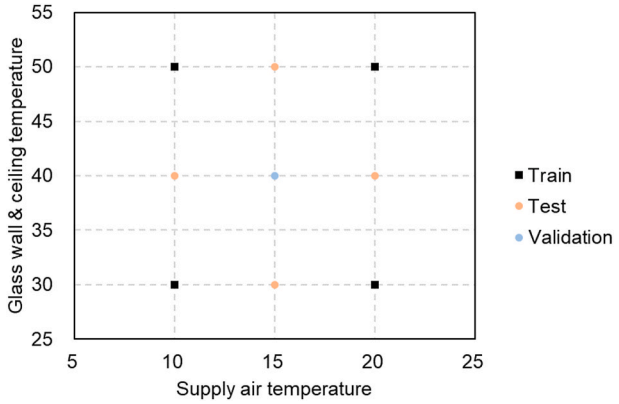


Fig. 14. Dataset for HKUST academic building.

(2) The POD-ANN significantly reduces the model complexity compared to using ANN directly. The scale of the neural network model parameters is merely about 1 % of the latter, demonstrating its exceptional lightweight nature, which could significantly reduce the computational resources (memory and computational time) required for indoor environment prediction, the calculation time for each run can be reduced by 63 %.
(3) The selection of training set data strongly influences the performance of these surrogate models. Increasing the quantity of training sets does not necessarily improve prediction accuracy. However, when utilizing the same number of training sets, the coverage of the entire data space within the training set leads to markedly different result.
(4) In comparison to traditional POD interpolation methods such as RBF, the application of ANN for the prediction of POD results exhibits enhanced extrapolation capabilities.
(5) The surrogate models examined in this article are engaged in a process of feature extraction and learning from CFD results before making predictions. It is clear from the principles of both POD and ANN that these surrogate models still possess certain limitations. The accuracy of the POD largely depends on the number of modes selected and the interpolation method used. On the other hand, ANN is a purely data-driven approach which lacks the capability to learn the physical knowledge inherent in fluid dynamics.

The current study has the following two limitations. First, the transferability of the method still needs further validation. Second, the focus of this study is on steady-state environmental conditions in building temperature distribution. However, POD will have more advantages in dynamic simulation theoretically. That is, the extracted flow field modes should be time dependent. This concept can be represented in Fig. 6 as the replacement of boundary conditions with timesteps. It is worth noting that while time series have an order, combinations of boundary conditions do not. Therefore, an important premise for this equivalence is that, according to the properties of matrices, swapping any two rows of a matrix does not change the result of the matrix's SVD decomposition. That is, shuffling the time series can also yield the same temporal modes.

Overall, these findings highlight the effectiveness of the POD-ANN surrogate model for rapid indoor environment prediction and emphasize the importance of careful training set selection in achieving accurate results. Furthermore, our study demonstrates the broad applicability of the POD-ANN under the context of built environment simulation.

The entire data, code and models are shared in Github (Link: https://github.com/LigeZhao/CFD-surrogate-models).

CRediT authorship contribution statement

Lige Zhao: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Qi Zhou:** Writing – review & editing, Validation, Supervision, Methodology, Data curation, Conceptualization. **Mengying Li:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Zhe Wang:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Data curation, Conceptualization.

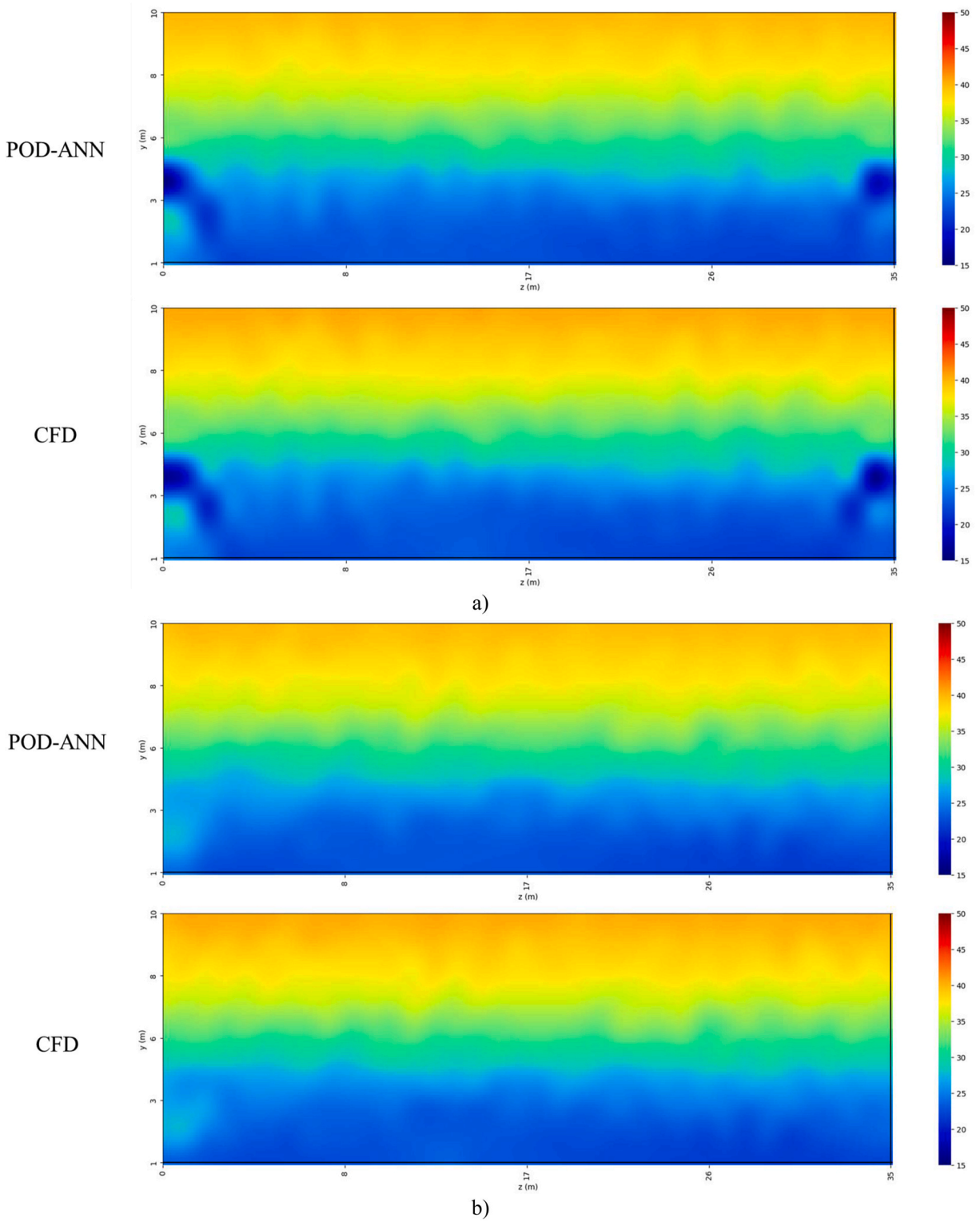


Fig. 15. Prediction of POD-ANN and CFD results: a) X = 6.5 m, b) X = 7.5 m

Table 9
Model accuracy of three models in HKUST academic building.

Criterion	Direct POD	Direct ANN (10 runs average)	POD-ANN (10 runs average)
R-square	0.96	0.99	0.99
MAE	0.58	0.29	0.25
RMSE	0.64	0.38	0.32
NRMSE	3.4 %	2.0 %	1.7 %

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

The work is substantially supported by the Project of Autonomous Cruise UVC Disinfection and Microclimate Air-conditioning Robot Topic#3 Thermal Management for the UVC LED Disinfection Robotics (FSUST21-SHCIRI07C) and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. C6003-22Y).

References

[1] J. Sundell, On the history of indoor air quality and health, Indoor Air 14 (s 7) (2004) 51–58.
[2] C. Huizenga, S. Abbaszadeh, L. Zagreus, E.A. Arens, Air quality and thermal comfort in office buildings: results of a large indoor environmental quality survey [Online]. Available: https://escholarship.org/content/qt7897g2f8/qt7897g2f8.pdf, 2006. (Accessed 22 November 2023).
[3] C. Li, S.-J. Yoo, K. Ito, Impact of indoor ventilation efficiency on acetone inhalation exposure concentration and tissue dose in respiratory tract, Build. Simulat. 16 (3) (Mar. 2023) 427–441, https://doi.org/10.1007/s12273-022-0954-4.
[4] C.A. Redlich, J. Sparer, M.R. Cullen, Sick-building syndrome, Lancet 349 (9057) (1997) 1013–1016.
[5] M. Deng, X. Wang, D. Li, C.C. Menassa, Digital ID framework for human-centric monitoring and control of smart buildings, Build. Simulat. 15 (10) (Oct. 2022) 1709–1728, https://doi.org/10.1007/s12273-022-0902-3.
[6] W.J. Fisk, H ealth and P roductivity G ains from B etter I ndoor E nvironments and their R elationship with B uilding E nergy E fficiency, Annu. Rev. Energy Environ. 25 (1) (Nov. 2000) 537–566, https://doi.org/10.1146/annurev.energy.25.1.537.
[7] W.J. Fisk, A.H. Rosenfeld, Estimates of improved productivity and health from better indoor environments, Indoor Air 7 (3) (Sep. 1997) 158–172, https://doi.org/10.1111/j.1600-0668.1997.t01-1-00002.x.
[8] L. Yang, H. Yan, J.C. Lam, Thermal comfort and building energy consumption implications–a review, Appl. Energy 115 (2014) 164–173.
[9] Q. Zhou, H. Zhong, L. Li, Z. Wang, AlphaMobileSensing: a virtual testbed for mobile environmental monitoring, Build. Simulat. 16 (7) (Jul. 2023) 1027–1040, https://doi.org/10.1007/s12273-023-1001-9.
[10] P.V. Nielsen, Fifty years of CFD for room air distribution, Build. Environ. 91 (2015) 78–90.
[11] Z. Zhang, W. Zhang, Z.J. Zhai, Q.Y. Chen, Evaluation of various turbulence models in predicting airflow and turbulence in enclosed environments by CFD: Part 2—comparison with experimental data from literature, HVAC R Res. 13 (6) (Nov. 2007) 871–886, https://doi.org/10.1080/10789669.2007.10391460.
[12] A. Stamou, I. Katsiris, Verification of a CFD model for indoor airflow and heat transfer, Build. Environ. 41 (9) (2006) 1171–1181.
[13] Y. Tominaga, T. Stathopoulos, CFD simulation of near-field pollutant dispersion in the urban environment: a review of current modeling techniques, Atmos. Environ. 79 (2013) 716–730.
[14] S. Gilani, H. Montazeri, B. Blocken, CFD simulation of stratified indoor environment in displacement ventilation: validation and sensitivity analysis, Build. Environ. 95 (2016) 299–313.
[15] T. van Hooff, B. Blocken, CFD evaluation of natural ventilation of indoor environments by the concentration decay method: CO2 gas dispersion from a semi-enclosed stadium, Build. Environ. 61 (2013) 1–17.
[16] C. Yang, P. Demokritou, Q. Chen, J. Spengler, Experimental validation of a computational fluid dynamics model for IAQ applications in ice rink arenas, Indoor Air 11 (2) (2001) 120–126.
[17] T. Hayashi, Y. Ishizu, S. Kato, S. Murakami, CFD analysis on characteristics of contaminated indoor air ventilation and its application in the evaluation of the effects of contaminant inhalation by a human occupant, Build. Environ. 37 (3) (2002) 219–230.
[18] D.N. Sørensen, P.V. Nielsen, Quality control of computational fluid dynamics in indoor environments, Indoor Air 13 (1) (2003) 2–17.
[19] P.V. Nielsen, Computational fluid dynamics and room air movement, Indoor Air 14 (Supplement 7) (2004) 134–143.
[20] T. Van Hooff, B. Blocken, Coupled urban wind flow and indoor natural ventilation modelling on a high-resolution grid: a case study for the Amsterdam ArenA stadium, Environ. Model. Software 25 (1) (2010) 51–65.
[21] C.F. Gao, W.L. Lee, Evaluating the influence of openings configuration on natural ventilation performance of residential units in Hong Kong, Build. Environ. 46 (4) (2011) 961–969.
[22] M.Z.I. Bangalee, S.Y. Lin, J.J. Miau, Wind driven natural ventilation through multiple windows of a building: a computational approach, Energy Build. 45 (2012) 317–325.
[23] R. Ramponi, B. Blocken, CFD simulation of cross-ventilation for a generic isolated building: impact of computational parameters, Build. Environ. 53 (2012) 34–48.
[24] Q. Chen, Ventilation performance prediction for buildings: a method overview and recent applications, Build. Environ. 44 (4) (2009) 848–858.
[25] B. Blocken, Computational Fluid Dynamics for urban physics: importance, scales, possibilities, limitations and ten tips and tricks towards accurate and reliable simulations, Build. Environ. 91 (2015) 219–245.
[26] B. Blocken, 50 years of computational wind engineering: past, present and future, J. Wind Eng. Ind. Aerod. 129 (2014) 69–102.
[27] W. Tian, X. Han, W. Zuo, M.D. Sohn, Building energy simulation coupled with CFD for indoor environment: a critical review and recent applications, Energy Build. 165 (2018) 184–199.
[28] S.L. Brunton, B.R. Noack, P. Koumoutsakos, Machine learning for fluid mechanics, Annu. Rev. Fluid Mech. 52 (1) (Jan. 2020) 477–508, https://doi.org/10.1146/annurev-fluid-010719-060214.
[29] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[30] A. Sempey, C. Inard, C. Ghiaus, C. Allery, Fast simulation of temperature distribution in air conditioned rooms by using proper orthogonal decomposition, Build. Environ. 44 (2) (2009) 280–289.
[31] J. Sietsma, R.J. Dow, Creating artificial neural networks that generalize, Neural Network. 4 (1) (1991) 67–79.
[32] A.H. Neto, F.A.S. Fiorelli, Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption, Energy Build. 40 (12) (2008) 2169–2176.
[33] L. Zhou, Optimization of Ventilation System Design and Operation in Office Environment, PhD Thesis, Concordia University, 2007 [Online]. Available: https://spectrum.library.concordia.ca/id/eprint/975579/. (Accessed 22 November 2023).
[34] G. Calzolari, W. Liu, Deep learning to replace, improve, or aid CFD analysis in built environment applications: a review, Build. Environ. 206 (2021) 108315.
[35] Q. Zhou, R. Ooka, Neural network for indoor airflow prediction with CFD database, in: Journal of Physics: Conference Series, IOP Publishing, 2021 012154 [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/2069/1/012154/meta. (Accessed 22 November 2023).
[36] Q. Zhou, R. Ooka, Influence of data preprocessing on neural network performance for reproducing CFD simulations of non-isothermal indoor airflow distribution, Energy Build. 230 (2021) 110525.
[37] Q. Zhou, R. Ooka, Performance of neural network for indoor airflow prediction: sensitivity towards weight initialization, Energy Build. 246 (2021) 111106.
[38] K. Taira, et al., Modal analysis of fluid flows: an overview, AIAA J. 55 (12) (Dec. 2017) 4013–4041, https://doi.org/10.2514/1.J056060.
[39] K. Taira, et al., Modal analysis of fluid flows: applications and outlook, AIAA J. 58 (3) (Mar. 2020) 998–1022, https://doi.org/10.2514/1.J058462.
[40] S.L. Brunton, B.R. Noack, Closed-loop turbulence control: progress and challenges, Appl. Mech. Rev. 67 (5) (2015) 050801.
[41] C.W. Rowley, S.T.M. Dawson, Model reduction for flow analysis and control, Annu. Rev. Fluid Mech. 49 (1) (Jan. 2017) 387–417, https://doi.org/10.1146/annurev-fluid-010816-060042.
[42] J.L. Lumley, Stochastic tools in turbulence, Courier Corporation [Online]. Available, 2007, pp. 19–25, https://books.google.com/books?hl=zh-CN&lr=&id=CsVlooboM3cC&oi=fnd&pg=PP1&dq=Stochastic+Tools+in+Turbulence+1970&ots=apG4r_PKix&sig=Wp1Duw6e7xr78qAOzUw3BtkrQFM (Accessed 22 November 2023).chapter 2.
[43] K. Taira, T. Colonius, The immersed boundary method: a projection approach, J. Comput. Phys. 225 (2) (2007) 2118–2137.
[44] L. Sirovich, Turbulence and the dynamics of coherent structures. I. Coherent structures, Q. Appl. Math. 45 (3) (1987) 561–571.
[45] K. Li, H. Su, J. Chu, C. Xu, A fast-POD model for simulation and control of indoor thermal environment of buildings, Build. Environ. 60 (2013) 150–157.
[46] K. Li, W. Xue, C. Xu, H. Su, Optimization of ventilation system operation in office environment using POD model reduction and genetic algorithm, Energy Build. 67 (2013) 34–43.
[47] K. Li, Z. Sha, W. Xue, X. Chen, H. Mao, G. Tan, A fast modeling and optimization scheme for greenhouse environmental system using proper orthogonal decomposition and multi-objective genetic algorithm, Comput. Electron. Agric. 168 (2020) 105096.
[48] K. Li, W. Zheng, W. Xue, Z. Wang, Fast reconstruction of indoor temperature field for large-space building based on limited sensors: an experimental study, Energy Build. 298 (2023) 113493.
[49] Y. Luo, Y. Song, Z. Tian, J. Fan, L. Zhang, Fast and accurate prediction of air temperature and velocity field in non-uniform indoor environment under complex boundaries, Build. Environ. 230 (2023) 109987.
[50] Y. Wang, B. Yu, Z. Cao, W. Zou, G. Yu, A comparative study of POD interpolation and POD projection methods for fast and accurate prediction of heat transfer problems, Int. J. Heat Mass Tran. 55 (17–18) (2012) 4827–4836.
[51] Q.I.U. Yasong, B.A.I. Junqiang, H.U.A. Jun, Flow field estimation method based on proper orthogonal decomposition and surrogate model, Hangkong XuebaoActa Aeronaut. Astronaut. Sin. 34 (6) (2013) 1249–1260.
[52] X. Jia, C. Gong, C. Li, Fast flow simulation method based on POD and BPNN, Xibei Gongye Daxue XuebaoJournal Northwest. Polytech. Univ. 39 (6) (2021) 1212–1221.
[53] H. Dongmei, H. Shiqing, H. Xuhui, Z. Xue, Prediction of wind loads on high-rise building using a BP neural network combined with POD, J. Wind Eng. Ind. Aerod. 170 (2017) 1–17.
[54] M. Wang, Q. Chen, Assessment of various turbulence models for transitional flows in an enclosed environment (RP-1271), HVAC R Res. 15 (6) (Nov. 2009) 1099–1119, https://doi.org/10.1080/10789669.2009.10390881.
[55] A. Standard, Thermal environmental conditions for human occupancy, ANSIASHRAE 55 5 (1992) [Online]. Available: https://cir.nii.ac.jp/crid/1574231875437564672. (Accessed 22 November 2023).
[56] Z. Wang, T. Hong, H. Li, M.A. Piette, Predicting city-scale daily electricity consumption using data-driven models, Adv. Appl. Energy 2 (2021) 100025.