Journal of Building Performance Simulation

ISSN: 1940-1493 (Print) 1940-1507 (Online) Journal homepage: www.tandfonline.com/journals/tbps20

Beyond CFD: explainable machine learning for efficient assessment of urban morphology impacts on pedestrian level wind and thermal environment

Xue Cui, Yu Li & Pengyuan Shen

To cite this article: Xue Cui, Yu Li & Pengyuan Shen (2026) Beyond CFD: explainable machine learning for efficient assessment of urban morphology impacts on pedestrian level wind and thermal environment, Journal of Building Performance Simulation, 19:2, 165-180, DOI: 10.1080/19401493.2025.2508500

To link to this article: https://doi.org/10.1080/19401493.2025.2508500

Published online: 28 May 2025.

JOURNAL OF BUILDING PERFORMANCE SIMULATION
2026, VOL. 19, NO. 2, 165–180
https://doi.org/10.1080/19401493.2025.2508500

METHOD

Beyond CFD: explainable machine learning for efficient assessment of urban morphology impacts on pedestrian level wind and thermal environment

Xue Cui[a], Yu Li[b,c] and Pengyuan Shen[d]

[a]Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China; [b]School of Architecture and Design, Harbin Institute of Technology, Harbin, People's Republic of China; [c]Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, Harbin, People's Republic of China; [d]Institute of Future Human Habitats, Shenzhen International Graduate School, Tsinghua University, Shenzhen, People's Republic of China

ABSTRACT

Urban morphology plays a crucial role in affecting pedestrian level wind velocity and air temperature. Most studies use CFD to simulate these parameters, which requires substantial computational resources, detailed inputs, and modeller expertise. To assess the impact of urban morphology on these parameters, comparative analyses are often used but are highly dependent on pre-designed scenarios. The study established two XGBoost-based models to predict wind velocity and air temperature efficiently while applying SHAP to investigate the effect of urban morphology. SHAP results revealed that the horizontal and vertical distances from the array centre showed a greater influence on microclimate parameters than conventional building design variables. This study introduces efficient machine learning models predicting pedestrian-level wind and thermal environment with high temporal granularity while providing quantifiable insights into the relationships between urban morphology variables and microclimate conditions through SHAP-based analysis, offering practical design optimization support for creating more comfortable urban environments.

ARTICLE HISTORY
Received 17 February 2025
Accepted 13 May 2025

KEYWORDS
Wind environment; air temperature; urban morphology; thermal comfort; computational fluid dynamics; machine learning

1. Introduction

In urban outdoor spaces, the wind and thermal environment at the pedestrian level play a crucial role in human comfort. Among the key influential factors, urban morphology significantly impacts pedestrian level wind velocity and air temperature (Dong et al. 2025; Shen et al. 2024; Sun et al. 2022). Extensive research has been conducted in this field (Du, Mak, and Li 2019; Weerasuriya et al. 2020; Wu, Zhan, and Quan 2021b; Zhang, Cui, and Song 2020), with two primary research priorities: predicting wind and thermal environment parameters and identifying the effects of urban morphology on these parameters.

Existing studies predominantly employed experimental and numerical approaches to predict wind and thermal environment parameters (Yang et al. 2023). The experimental approach relies on field measurements and wind tunnel tests, while the numerical approach includes computational fluid dynamics (CFD) and data-driven models (Yang et al. 2023). Compared to experimental approaches, numerical approaches offer flexibility in modifying parameters, simulating multiple scenarios, and conducting comparative analyses (Zheng, Chen, and Yang 2023). Among numerical approaches, CFD was widely used to assess pedestrian level wind and thermal environments (Mochida and Lun 2008; Shen and Wang 2020; Wang et al. 2025). However, CFD simulations are highly computationally demanding, often requiring several hours to multiple days to complete (Liu and Niu 2019; Wu, Zhan, and Quan 2021a). Additionally, they demand detailed input parameters and specialized expertise from modellers (Kastner and Dogan 2023), making them challenging for early-stage design assessments. In contrast, data-driven models offer a promising alternative by enabling fast and accurate predictions of wind and thermal parameters. These models utilize machine learning algorithms to develop predictive models that capture relationships between known urban morphology inputs and corresponding wind and thermal outputs, enabling the prediction of unknown wind and thermal parameters in other unseen urban morphologies. Yang et al. (2023) identified four primary data sources for developing machine learning models in this field: weather station records, field measurements, wind tunnel tests, and CFD simulations. In addition, machine learning models in this field demonstrate high predictive accuracy, with coefficients of determination (R^2) typically ranging from 0.8 to

CONTACT Pengyuan Shen pengyuan_pub@163.com

0.9 (Yang et al. 2023). For example, Giljarhus and Hågbo (2024) developed a regression model based on CFD-simulated data to predict pedestrian level wind velocity at the building scale, achieving R^2 values ranging from 0.810 to 0.991. Similarly, Mokhtar et al. (2021) and Clemente et al. (2024) introduced similar deep learning-based workflows to estimate pedestrian level wind-related factors in urban environments. Notably, machine learning models greatly enhanced computational efficiency, with studies showing a reduction in computational time from approximately eight hours for CFD simulations to just approximately four seconds for data-driven models (Kastner and Dogan 2023). These advances highlight the need for further exploring the integration of machine learning techniques in urban climate research.

When assessing the effect of urban morphology on pedestrian level wind and thermal environment, most studies used comparative analysis to explore their differences under different urban morphologies. For instance, Li et al. (2020) used CFD to simulate outdoor wind velocity and air temperature of five building array scenarios, analyzing the effect of frontal area density of the building array on pedestrian level thermal comfort. Such studies commonly simulated wind and thermal parameters across various urban morphology scenarios individually, followed by statistical analyses to elucidate the effect of urban morphology on wind and thermal environment. It highly relies on pre-designed scenarios and is still time-consuming. In addition, other studies used an overall coefficient to determine the impact of urban morphology on pedestrian level wind and thermal environment. For instance, Zhang, Cui, and Song (2020) used R^2 to explore the relationship between urban morphology and outdoor wind and thermal comfort. However, this global coefficient makes it difficult to interpret both the underlying reasons and the impact of urban morphology on wind and thermal environment in specific scenarios. As an advanced explanatory technique to interpret black-box machine learning models, SHapley Additive exPlanations (SHAP) can well quantify the contribution of input features on model outputs by SHAP values from global to local levels (Lundberg et al. 2020). It has been widely used in explaining the prediction behaviour of machine learning models and plays a crucial role in the interpretable analysis field (Mi, Li, and Zhou 2020). Some studies have applied SHAP at the room level to investigate the effects of specific variables (e.g. indoor and outdoor environmental parameters) on thermal comfort (Lan, Hou, and Gou 2023; Yang et al. 2022), while others have applied SHAP at the building level to investigate the influence of household characteristics on building energy use (Cui et al. 2024). However, few studies have applied SHAP at the urban level to explore the relationship between urban morphology and pedestrian level wind and thermal environment. Since wind velocity and air temperature are commonly regarded as key parameters for representing the pedestrian level wind and thermal environment, as well as essential variables for calculating wind and thermal comfort (Du et al. 2017; Jeong et al. 2022), it is necessary to first develop machine learning models that can efficiently and accurately predict pedestrian level wind velocity and air temperature. Subsequently, SHAP can be applied to these models to interpret the contributions of input urban morphology design variables to predicted wind velocity and air temperature outputs.

In summary, existing studies predicting pedestrian level wind and thermal environment and investigating the effects of urban morphology on them have entailed two research gaps. Firstly, the simulation approach is challenging for assessing wind and thermal environment due to the high requirements of computational time and modeller knowledge and expertise. Secondly, comparative analysis and global coefficients are limited in explaining the effects of urban morphology on wind and thermal environment parameters, due to constraints related to scenario diversity, time limitation, and interpretability. To address these two gaps, the study aims to (1) develop machine learning models to quickly and accurately predict pedestrian level wind velocity and air temperature; (2) apply SHAP to the established machine learning prediction models to investigate the effect of urban morphology on pedestrian level wind velocity and air temperature in an explainable way. To achieve these two objectives, the study was organized into three main sections: (1) establish databases, including urban morphology-related design variables and the corresponding pedestrian level wind velocities and air temperatures; (2) develop two separate machine learning models for predicting pedestrian level wind velocity and air temperature; (3) apply SHAP to each prediction model to investigate the effect of urban morphology on wind velocity and air temperature. The novelty of this study lies in the development of two predictive models capable of estimating pedestrian level wind velocity and air temperature within seconds, thereby eliminating the need for time-consuming simulations. In addition, the SHAP-based effect analysis results provide valuable policy implications for early-stage urban planning and building layout design, contributing to optimizing the urban microclimate and enhancing the efficiency of the decision-making process.

2. Methodology

Figure 1 proposed a research framework for establishing wind velocity and air temperature prediction models

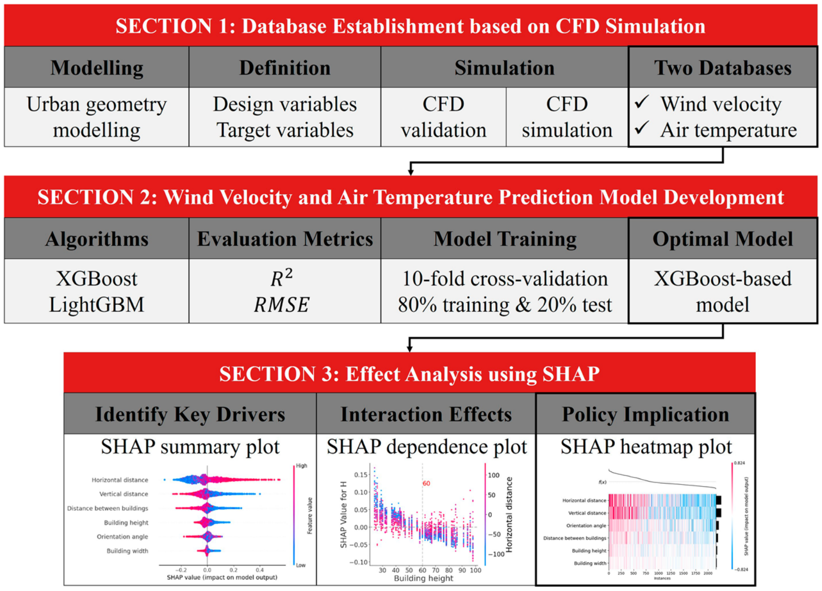


Figure 1. Research framework for establishing wind velocity and air temperature prediction models and SHAP-based effect analysis.

and conducting SHAP-based effect analysis. It comprises three main sections: (1) establishing two databases based on CFD simulations, which includes urban geometry modelling, defining design and target variables, CFD validation and simulation; (2) developing wind velocity and air temperature prediction models using the two established databases, involving determining machine learning algorithms and evaluation metrics, model training, and identification of the optimal model; and (3) conducting SHAP-based effect analysis to identify key drivers influencing pedestrian level wind velocity and air temperature based on SHAP summary plot, to explore how the effect of one urban morphology design variable on wind velocity and air temperature interacts with another based on SHAP dependence plot, and to derive policy implications for the early stages of urban planning and building layout design based on SHAP heatmap plot. The results are not only applicable to urban morphologies similar to those examined in this study, but the research methodology and framework are also generalizable and can be applied to other case studies with different urban morphologies.

2.1. Database establishment based on CFD simulation

CFD simulations using Ansys Fluent 2019 R3 were conducted to calculate pedestrian level wind velocity and air temperature under various urban morphologies for data collection purposes. The study first modelled a typical urban geometry representative of common urban morphologies. Urban morphology design variables, along with predicted wind velocity and air temperature variables, were then defined. The ranges and discretization of these design variables were determined for Latin Hypercube Sampling (LHS), enabling the generation of diverse urban morphology scenarios. Before the actual CFD simulations, a CFD validation was performed to select the appropriate mesh scheme and turbulence model. Finally, CFD simulations were carried out for each sampled urban morphology, and databases were established that link the urban morphology design variables to their corresponding wind velocity and air temperature values.

2.1.1. Urban geometry modelling

The study modelled a building array consisting of nine cuboids to represent common urban morphologies (Figure 2). This configuration was based on Case C proposed by the Working Group of the Architectural Institute of Japan (AIJ) (Tominaga et al. 2008) and has been widely used in studies concentrating on outdoor wind and thermal environments (Chen, Rong, and Zhang 2021; Gromke and Blocken 2015; Liu et al. 2019a; 2019b; Liu and Niu 2019; Tong, Chen, and Malkawi 2016). As the study focused on pedestrian level wind and thermal environments, sidewalks surrounding the buildings were designed as test areas, indicated by the black-filled regions in Figure 2(b). Due to the considerable spatial variations in wind velocity and air temperature, 36 test areas were positioned around all four sides of the nine building cuboids. This arrangement enabled monitoring of spatial variations in wind velocity and air temperature. The sidewalks were designed to be five metres (m) away from the buildings, positioned 1.5 metres above ground

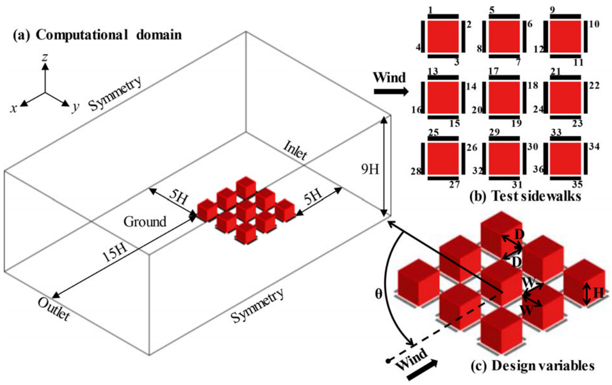


Figure 2. Description of urban geometry modelling.

level, with a width of five metres and a length matching the width of the buildings.

The computational domain for the CFD simulation was defined based on the building height (H), following the guidelines of the AIJ (Mochida et al. 2016). As shown in Figure 2(a), the distance from the inlet to the windward surface of the first building is 5H, while the distance from the side symmetry boundaries to the building sidewalls is also 5H. The outlet is positioned 15H downstream from the leeward surface of the last building, and the top symmetry boundary is set 8H above the building roofs.

2.1.2. Definition of design and target variables

The study selected building height (H), building width (W), distance between buildings (D), and orientation angle (θ) as design variables to control urban morphology since these variables significantly affect pedestrian level wind and thermal environments (Du, Mak, and Li 2019; Wu, Zhan, and Quan 2021b). To generate diverse urban morphologies, the ranges and discretization of these four design variables were defined, as shown in Table 1. Based on these parameters, LHS was employed to randomly generate unique design schemes in this study, as LHS can efficiently generate limited samples while ensuring better coverage of the high-dimensional design space (Wu, Zhan, and Quan 2021b). The ranges and discretization for H, W, and D were determined by referring to the 'Code for Fire Protection Design of Buildings GB50016-2014' (Ministry of Public Security of the People's Republic of China 2018) and existing literature (Du, Mak, and Li 2019). Although the same design scheme is theoretically symmetric at θ and 90°-θ, the variations in H, W, and D of the urban morphology design schemes generated through LHS result in asymmetry at θ and 90°-θ. Consequently, to ensure diverse urban morphologies, the range of θ was defined from 0° to 90° with a discretization of 1°. The area-weighted wind velocity (M_Vel) and air temperature

Table 1. Description of design variables.

Design variables	Range: Lower	Range: Upper	Discretization
H	24 m	99 m	1 m
W	20 m	50 m	1 m
D	24 m	99 m	1 m
θ	0°	90°	1°

(M_Ta) within each test sidewalk were selected as the target variables, with their definitions provided in Equations (1)–(2).

$$M\_Vel = \frac{1}{A}\int_A Vel dA \quad (1)$$

$$M\_Ta = \frac{1}{A}\int_A Ta dA \quad (2)$$

Where *Vel* and *Ta* represent the wind velocity and air temperature at any location within a test sidewalk, and *A* denotes the area of the test sidewalk (Du, Mak, and Li 2019).

2.1.3. CFD validation

Since the AIJ provides open-access wind tunnel results for Case C (Architectural Institute of Japan 2020), the study conducted CFD simulation for Case C using the detailed inflow profiles provided by AIJ. The wind tunnel results and CFD simulation outcomes for Case C were then compared for validation purposes. This process facilitated the selection of the appropriate generation method of the mesh scheme and turbulence model for subsequent CFD simulations. Figure 3(a) shows the geometry and test points of AIJ Case C, while Figure 3(b and c) present the corresponding inflow profiles. For validation, the study compared wind tunnel results with CFD simulation results at test points located at a height of 0.1H.

Three processes were conducted for the CFD validation. First, a mesh sensitivity test was performed on Case C using three mesh schemes, including coarse, medium, and fine, combined with the standard k-ϵ turbulence model. All three mesh schemes were configured with a growth rate of 1.2, 10 inflation layers, and a first-layer thickness of 0.0011 m. The global element sizes were 0.054 m for the coarse mesh, 0.036 m for the medium mesh, and 0.024 m for the fine mesh, while the element sizes around the buildings were 0.027, 0.018, and 0.012 m, respectively. After identifying the optimal mesh scheme, the study applied it to two additional turbulence models, including the RNG k-ϵ and realizable k-ϵ model, for further Case C simulations. Finally, the simulation results from the three turbulence models, using the selected mesh scheme, were compared with wind tunnel experimental results, allowing the study to identify the most

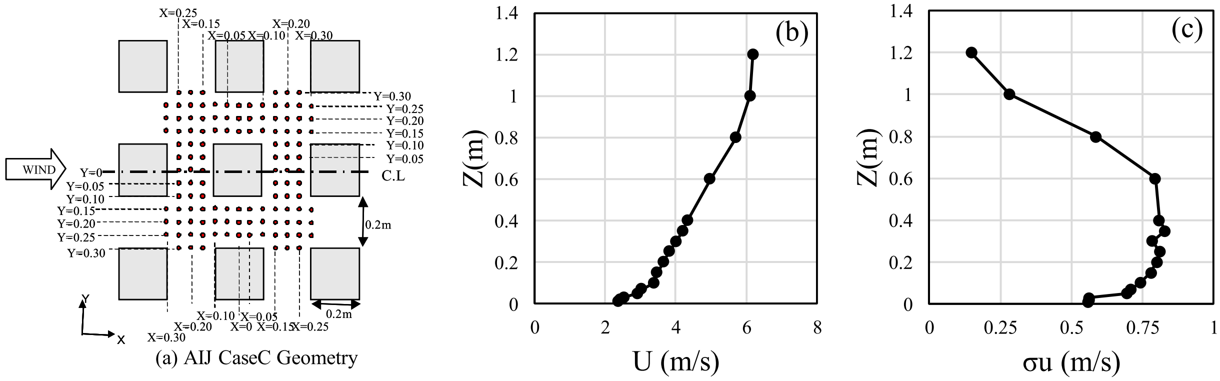


Figure 3. Description of AIJ Case C.

Table 2. *R* values between the Case C wind tunnel results and the simulation results using the medium mesh scheme.

Turbulence model	Mesh system	*R*
Standard k-ϵ	***Medium***	***0.84***
RNG k-ϵ	Medium	0.55
Realizable k-ϵ	Medium	0.83

suitable turbulence model and assess the consistency between the Case C simulations and wind tunnel results. Notably, the coupled and second-order upwind schemes were applied as the solution methods in all simulations throughout this process.

Points along eight lines of $x = -0.25$, $x = -0.2$, $x = 0.2$, $x = 0.25$, $y = -0.25$, $y = -0.2$, $y = 0.2$, and $y = 0.25$, at a height of 0.1H, were selected as target points for the mesh sensitivity test. Figure 4 presents the comparison results of the three mesh schemes, where the velocity ratio, calculated by normalizing the wind velocity with the inflow velocity at the same height, served as the comparison criterion. Overall, the medium and fine mesh schemes produced closely aligned results, both differing significantly from those under the coarse mesh. Given its comparable accuracy to the fine mesh but lower computational cost, the medium mesh scheme was selected as the appropriate mesh. The study further employed the Spearman Correlation Coefficient (*R*) to further assess the correlation between Case C wind tunnel results and simulation outcomes across the three turbulence models based on the selected medium mesh. As shown in Table 2, *R* values were 0.84 for the standard k-ϵ model, 0.55 for the RNG k-ϵ model, and 0.83 for the realizable k-ϵ model. The standard k-ϵ model exhibited a significantly higher *R* than the RNG k-ϵ model and a slightly higher *R* than the realizable k-ϵ model, making it the most suitable turbulence model. Therefore, combining the medium mesh scheme with the standard k-ϵ model was determined to be the optimal configuration for simulating Case C, yielding the best agreement with the wind tunnel results.

2.1.4. CFD simulation

Beijing was selected as the study area, with the typical urban geometry of AIJ Case C applied to Beijing's weather conditions for simulation. As a highly urbanized city with diverse and complex building structures (Zhang et al. 2022), studying the effect of its urban morphology on pedestrian level wind velocity and air temperature is crucial. This study focused on winter wind and thermal environments, as Beijing is located in the cold climate zone of China, where pedestrians face significant outdoor environmental challenges. Investigating the effect of urban morphology under pedestrian level winter conditions can help enhance outdoor wind and thermal comfort in Beijing.

Based on the CFD validation results, the study utilized the standard k-ϵ model for wind environment simulations. For the k-ϵ model, the vertical inflow wind velocity$U(z)$, turbulent kinetic energy $k(z)$, and turbulence dissipation rate $\varepsilon(z)$ were defined according to the Equations (3)–(6) proposed by AIJ (Tominaga et al. 2008):

$$U(z) = U_s\left(\frac{z}{z_s}\right)^{\alpha} \tag{3}$$

$$I(z) = \frac{\sigma_u(z)}{U(z)} = 0.1\left(\frac{z}{z_G}\right)^{(-\alpha-0.05)} \tag{4}$$

$$k(z) = \frac{\sigma_u^2(z) + \sigma_v^2(z) + \sigma_w^2(z)}{2} \cong \sigma_u^2(z) = (I(z)U(z))^2 \tag{5}$$

$$\varepsilon(z) = C_\mu^{1/2}k(z)\frac{U_s}{z_s}\alpha\left(\frac{z}{z_s}\right)^{(\alpha-1)} \tag{6}$$

Where U_s is the wind velocity at the reference height z_s, α is the power-law exponent determined by terrain category, $I(z)$ is the turbulent intensity, z_G is the atmospheric boundary layer height, and C_μ is the turbulent viscosity constant that is 0.09.

To capture the air temperatures of the test sidewalks, the study employed the Energy model and the Discrete

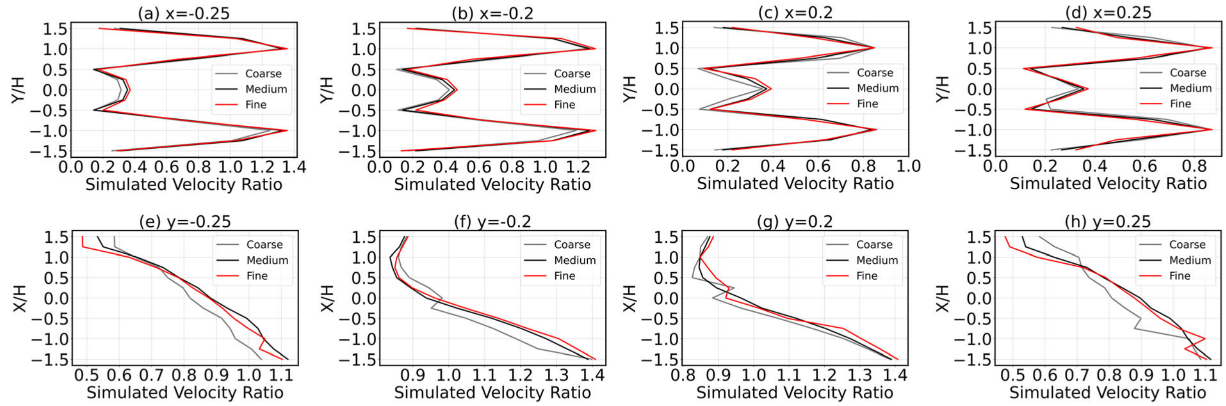


Figure 4. Comparison of velocity ratios across the three mesh schemes used in the CFD validation.

Ordinates (DO) radiation model to simulate thermal environments as they have demonstrated high accuracy in prior research (Li et al. 2020). First, the Solar Calculator was used to determine the sun position based on inputs such as specific date and time, longitude, and latitude. Then, the Ray-Tracing model offered the solar radiation incident on exposed surfaces, which was applied as a boundary condition and added into the energy equation. Finally, the DO model was used to yield the radiant heat fluxes between surfaces. Material properties were defined following previous literature (Li et al. 2020): the building material has a density of 2400 kg/m^3, specific heat of 750 J/kg K, thermal conductivity of 1.7 W/m K, absorption coefficient of 0.9, scattering coefficient of 0, refractive index of 1.7, and emissivity of 0.7; the ground material has a density of 2360 kg/m^3, specific heat of 920 J/kg K, thermal conductivity of 0.75 W/m K, absorption coefficient of 0.9, scattering coefficient of −10, refractive index of 1.92, and emissivity of 0.95. The heat transfer coefficients (h_c) of the ground and building walls were calculated by the empirical equation (Yang et al. 2017):

$$h_c = 5.7 + 3.8V_{air} \tag{7}$$

Where V_{air} is the airflow velocity.

Winter in Beijing typically spans from December to February. To determine the specific simulation time, the study calculated the mean Universal Thermal Climate Index (UTCI) for each hour of winter daytime based on the local EPW file. As shown in Figure 5, the mean UTCI reaches its lowest point at 9am, coinciding with the morning rush hour in Beijing (Liu et al. 2021). This indicates that pedestrians experience the most challenging environmental conditions during their commute at this peak hour. Therefore, 9am was selected as the simulation time. The study determined the simulation inflow parameters by calculating the mean wind velocity, air temperature, and solar radiation-related parameters, as well as the mode of wind direction at 9am in winter based on the local EPW file. The simulation date was selected based on the typical winter period with the nearest average temperature in the EPW file. Additionally, the power-law exponent and atmospheric boundary layer height were set following the 'Standard for Green Performance Calculation of Civil Buildings in China' (Ministry of Housing and Urban-Rural Development of the People's Republic of China 2018). Table 3 summarizes the CFD simulation parameters and settings.

The simulations were conducted on a high-performance computer with a 12th Gen Intel® Core™ i7-12700 K (3.60 GHz) processor and 32 GB RAM, utilizing parallel processing. The computational domain setup, mesh generation strategy, turbulence model choice, and solution method followed the same configurations used in the CFD validation described in section 2.1.3. Specifically, the standard k-ϵ turbulence model was employed

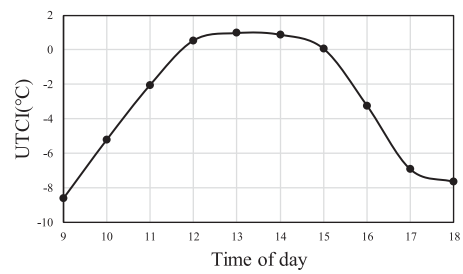


Figure 5. Mean UTCI in winter daytime.

Table 3. CFD simulation parameters and settings.

Parameters	Settings
Global location	Longitude = 116.28°E, Latitude = 39.93°N, Altitude = 55 m, Timezone = GMT+8
Date and time	9am, 14th December
Inlet	α = 0.22, z_s = 10 m, U_s = 2.69 m/s, z_G = 400 m, wind direction = 0°, air temperature = −3.81°C
Ground and building wall	No-slip stationary wall
Lateral and top boundary	Symmetry
Outlet	Outflow

with the medium mesh scheme, which had achieved an R value of 0.84 in the CFD validation, along with a coupled and second-order upwind solution scheme for the CFD simulations. Notably, only the mesh generation approach for the CFD simulations remained consistent with the validation case, however, the specific mesh sizes differed due to variations in the inflow profiles. Therefore, a mesh sensitivity test was performed again for the CFD simulations, and similarly, three mesh schemes were generated, including coarse, medium, and fine. All three mesh schemes were configured with a growth rate of 1.2, 10 inflation layers, and a first-layer thickness of 0.027 m. The global element sizes for coarse, medium, and fine meshes were 39, 27, and 21 m, respectively, while the element sizes around buildings were 6.5, 4.5, and 3.5 m. The study first calculated the R values between the results obtained from the fine mesh and those from the coarse and medium meshes to compare three results. In addition, the Grid Convergence Index (GCI), which quantifies the grid convergence error based on Richardson's extrapolation method (Roache 1997), was adopted to select the optimal mesh scheme. The key equations used for calculating the GCI were presented below (Celik et al. 2008):

The apparent order p of the convergence is calculated by:

$$p = \frac{1}{\ln(r_{21})}\left|\ln\left|\frac{f_3 - f_2}{f_2 - f_1}\right| + \ln\left(\frac{r_{21}^p - s}{r_{32}^p - s}\right)\right| \tag{8}$$

Where $s = \mathrm{sgn}\left(\frac{f_3 - f_2}{f_2 - f_1}\right)$, f_1, f_2, f_3 are the solutions on the fine, medium, and coarse meshes, r_{21} is the refinement ratio between fine and medium meshes, r_{32} is the refinement ratio between medium and coarse meshes.

The extrapolated value is calculated by:

$$f_{ext}^{21} = \frac{r_{21}^p f_1 - f_2}{r_{21}^p - 1} \tag{9}$$

The GCI between fine and medium meshes is calculated by:

$$GCI_{21} = \frac{1.25|e_a^{21}|}{r_{21}^p - 1} \tag{10}$$

The GCI between medium and coarse meshes is calculated by:

$$GCI_{32} = \frac{1.25|e_a^{32}|}{r_{32}^p - 1} \tag{11}$$

Where $e_a^{21} = \frac{f_1 - f_2}{f_1}$, $e_a^{32} = \frac{f_2 - f_3}{f_2}$, which are the approximate relative error.

Table 4 summarizes the total number of mesh elements, simulation runtimes, and R values with the fine mesh. The high correlations ($R \geq 0.96$) between the coarse and medium meshes compared to the fine mesh indicated strong agreement, with the medium mesh showing slightly higher R values than the coarse mesh. The average GCI for the 36 test sidewalks was 5.06% for the velocity ratio between the fine and medium mesh schemes, and 5.09% between the medium and coarse mesh schemes. Meanwhile, the average GCI for the air temperature ratio was 0.46% between the fine and medium mesh schemes, and 0.40% between the medium and coarse mesh schemes. Overall, all GCI values indicated satisfactory results, with the fine mesh scheme offering no substantial improvement over the medium mesh scheme. Therefore, the medium mesh was ultimately selected as the optimal mesh scheme for CFD simulations, balancing accurate simulations of wind velocity and air temperature with a manageable number of mesh elements and reasonable simulation time. Additionally, Figure 6 compares the velocity ratios and air temperature ratios across the three mesh schemes, showing that while the results are generally similar, the medium mesh aligns more closely with the fine mesh across most sidewalks.

Based on the previously defined design and target variables, 60 urban morphology design schemes were randomly generated by LHS. Figure 7 shows some of the sampled geometries. All 60 design schemes adhered to the simulation parameters and settings outlined in Table 3, utilizing the medium mesh. Each design scheme included 36 sidewalk test areas. The study incorporated the horizontal distance (Horiz_D) and vertical distance (Verti_D) between each sidewalk centre and the centre of the building array as additional design variables (Figure 8). The directionality of these distances was defined based on the relative position of the sidewalk to the building array centre. A total of 2,160 (60∗36) sidewalk samples were generated, each containing design variable data, horizontal and vertical distance information, and corresponding wind velocity and air temperature values. This dataset enables a comprehensive analysis of the effect of urban morphology on wind velocity and air temperature while capturing its spatial distribution variations.

2.2. Prediction model development based on machine learning

The study selected tree-based algorithms to develop the machine learning model for predicting pedestrian level wind velocity and air temperature. eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) were selected due to their superior performance in previous studies (Liu et al. 2022;

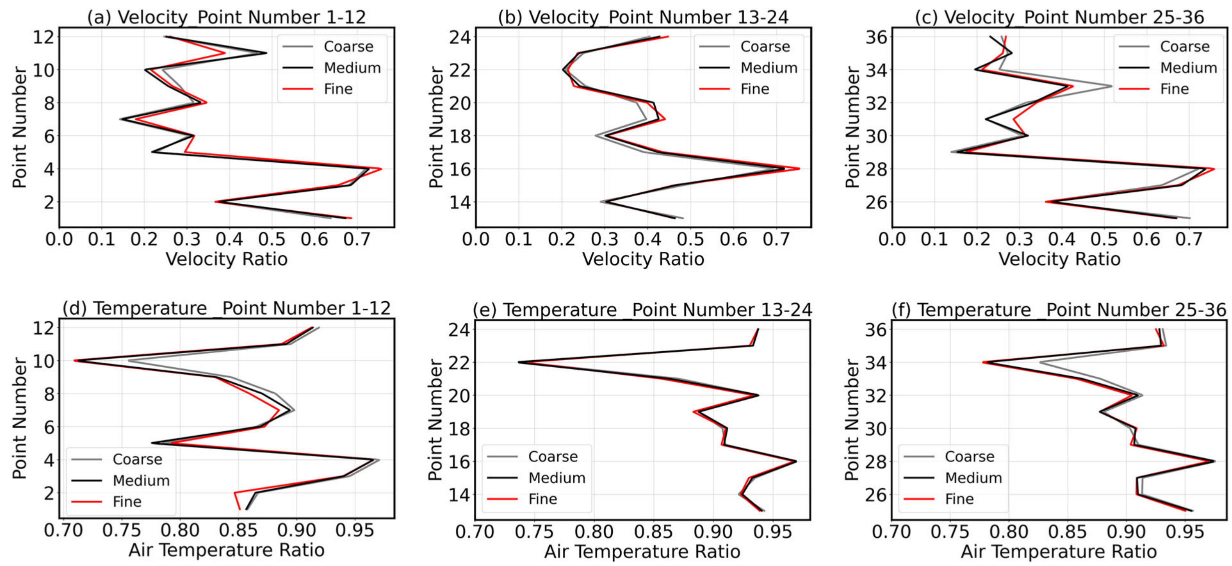


Figure 6. Comparison of velocity ratios and air temperature ratios across the three mesh schemes used in the CFD simulations.

Table 4. The number of mesh elements, simulation runtimes, and R values between coarse, medium, and fine mesh schemes.

			R with the fine mesh	
Mesh	Number of mesh elements	Runtimes	Wind velocity	Air temperature
Coarse	574, 371	1.40h	0.96	0.98
Medium	***1,768,283***	***3.97h***	***0.97***	***0.99***
Fine	3, 223, 059	7.07h	1.00	1.00

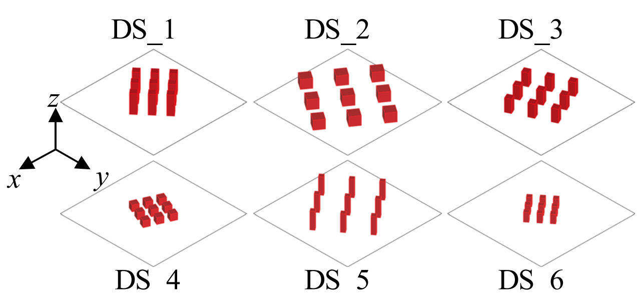


Figure 7. Geometries of some design schemes sampled with LHS.

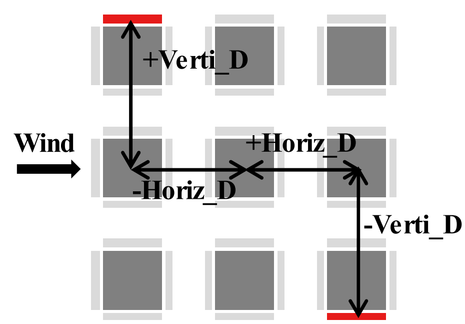


Figure 8. Horizontal and vertical distance.

Moon, Rho, and Baik 2022). The dataset from CFD simulations was split into 80% for training and 20% for test. Ten-fold cross-validation was conducted on the training set to optimize the model hyperparameters. The model predictive capability was first assessed on test data to analyze its generalization to unseen data. The model was then assessed on the entire dataset since it provided more comprehensive information than the training or test sets alone, which ensured that the SHAP-based effect analysis results were more robust and informative.

The Coefficient of Determination (R^2) and Root Mean Square Error (RMSE) were used to evaluate the model performance. R^2 represents the goodness of fit of models (Zhang et al. 2023), while RMSE measures the mean deviation between predicted and actual values (Liu et al. 2022). A higher R^2 and lower RMSE indicate better model performance. The calculations of these two metrics are presented in Equations (12)–(13):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_{Prediction,i} - y_{Actual,i})^2}{\sum_{i=1}^{n} (y_{Actual,i} - \bar{y}_{Actual})^2} \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_{Prediction,i} - y_{Actual,i})^2} \quad (13)$$

Where n represents the total number of samples in the dataset, $y_{Prediction,i}$ and $y_{Actual,i}$ denote the predicted and actual values of the i^{th} sample, and $\bar{y}_{Actual}$ is the mean of actual values.

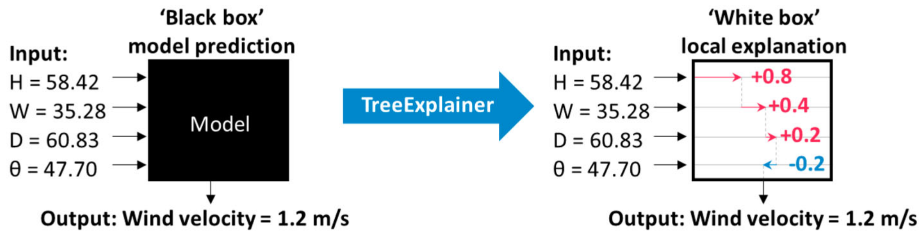


Figure 9. Graphical illustration of SHAP local explanation, adapted from Lundberg et al. (2020).

2.3. Effects analysis based on SHAP

Machine learning models are usually considered black-box models, making it challenging to interpret their model predictions. The Shapley value interprets the contribution of each input to model output based on its marginal impact (Shapley 1953). Equation (14) displays the Shapley value $\emptyset_i$ calculation (Lundberg and Lee 2017):

$$\emptyset_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)] \tag{14}$$

Here, F is the set including all features, S is the subset without the i^{th} feature, and $f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S)$ is the prediction difference, including and excluding the i^{th} feature.

Lundberg and Lee proposed SHAP in 2017, which retained all the benefits of Shapley value while significantly improving its computational efficiency, reducing the complexity from exponential to polynomial level (Lundberg et al. 2020). As a result, SHAP can provide local explanations for individual sample predictions while simultaneously capturing the global structure of feature contributions (Lundberg et al. 2020), with SHAP values measuring the contribution of each feature. The study applied SHAP to the optimal wind velocity and air temperature prediction models to interpret the influence of urban morphology on wind velocity and air temperature from both global and local perspectives. Figure 9 presents the graphical illustration of SHAP local explanation, demonstrating how it interprets model prediction behaviour.

3. Result and discussion

3.1. Wind velocity and air temperature databases

Table 5 summarizes the simulation results for 2,160 sidewalk samples. Figure 10 illustrates the distribution of simulated air temperature and wind velocity across these samples. Compared to the inflow conditions of an air temperature of −3.81°C and a wind velocity of 2.69 m/s, urban morphology exhibited a more pronounced influence on the wind environment. This is evident as the median simulated wind velocity decreased to 0.98 m/s, while the median simulated air temperature showed a slight increase to −3.35°C.

Table 5. Statistical description of simulation results.

Variables	Mean	Standard deviation	Min	Max
H	58.42	20.33	24.00	98.00
W	35.28	9.29	20.00	50.00
D	60.83	22.62	24.00	98.00
θ	47.70	25.93	3.00	90.00
|Horiz_D|	68.27	47.71	0.00	180.50
|Verti_D|	68.27	47.71	0.00	180.50
M_Ta	−3.32	0.20	−3.71	−1.87
M_Vel	1.24	0.77	0.09	3.30

3.2. Wind velocity and air temperature prediction models

3.2.1. Selection of the optimal model

The model was first assessed on test data to assess prediction accuracy and on the entire data to support SHAP analysis. Additionally, the study documented the validation results of the developed models. Table 6 summarizes the performance of the wind velocity prediction models. Overall, both LightGBM-based and XGBoost-based models demonstrated satisfactory performance, achieving R^2 of approximately 0.80 on validation data, around 0.85 on test data, and around 0.96 on entire data, and RMSE around 0.35 m/s for validation data, around 0.29 m/s for test data, and around 0.16 m/s for the entire data. Notably, the XGBoost-based wind velocity prediction model outperformed the LightGBM-based model across validation, test, and entire data. Therefore, the XGBoost-based model was selected as the optimal pedestrian level wind velocity prediction model. Figure 11 illustrates the comparison between actual and predicted wind velocities for the test and entire data, showing that most samples closely align with the diagonal line, with only a few outliers.

Table 7 presents the performance of the air temperature prediction models. Both LightGBM-based and XGBoost-based models demonstrated satisfactory results, with R^2 of approximately 0.70 on validation data, around 0.75 on test data, and exceeding 0.90 on the entire data, and RMSE of 0.11°C for both validation and

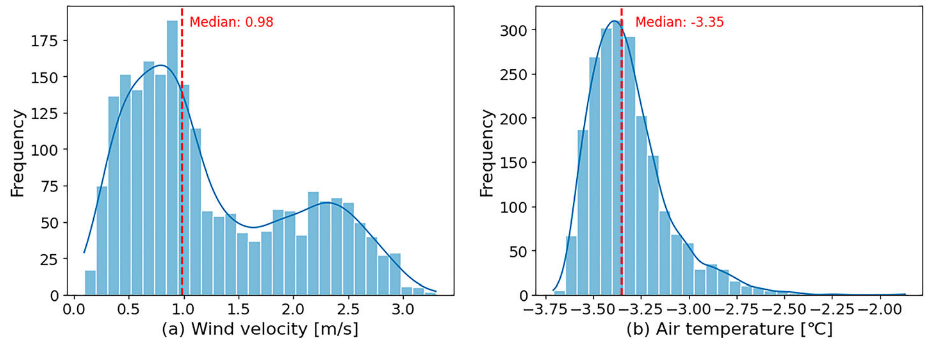
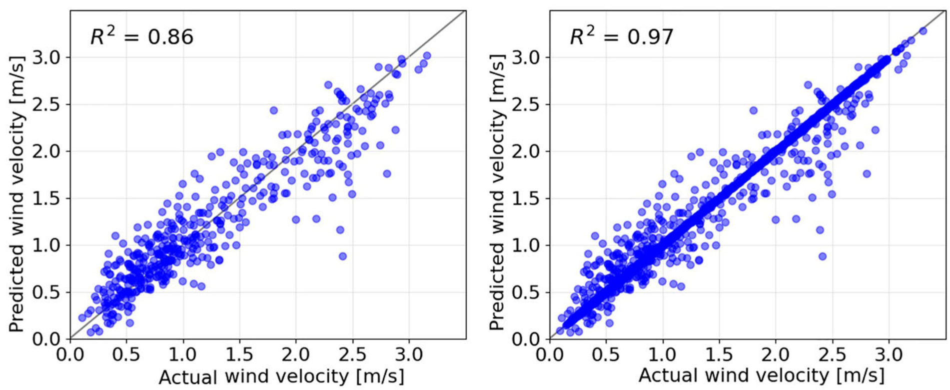


Figure 10. Distribution of target variables.



(a) On test data (b) On entire data

Figure 11. The actual versus the predicted values of XGBoost-based wind velocity model. (a) On test data (b) On entire data.

Table 6. Performance of wind velocity prediction models.

Dataset	Criteria	XGBoost	LightGBM
Validation data	R^2	***0.81***	0.79
	RMSE	***0.34***	0.35
Test data	R^2	***0.86***	0.85
	RMSE	***0.28***	0.29
Entire data	R^2	***0.97***	0.96
	RMSE	***0.13***	0.16

Table 7. Performance of air temperature prediction models.

Dataset	Criteria	XGBoost	LightGBM
Validation data	R^2	***0.70***	0.69
	RMSE	0.11	0.11
Test data	R^2	***0.75***	0.74
	RMSE	0.11	0.11
Entire data	R^2	***0.94***	0.91
	RMSE	***0.05***	0.06

test data, and approximately 0.05°C for the entire data. The XGBoost-based air temperature model slightly outperformed the LightGBM-based model across validation, test, and entire data. Consequently, the XGBoost-based model was selected as the optimal pedestrian level air temperature prediction model. Figure 12 illustrates the comparison between actual and predicted air temperatures for the test and entire data, showing high goodness of fit for most samples, with only a few deviations.

3.2.2. Reliability of the optimal model

To ensure the reliability of the selected optimal prediction models, the study compared their model performance with previous prediction models in a similar field. Given that RMSE values are unit-dependent, R^2 was chosen as the comparison criteria as its dimensionless nature allows for a more intuitive comparison. Table 8 summarizes the R^2 results from previous prediction models in the field of outdoor wind and thermal environment prediction, along with the results from this study.

Overall, R^2 values varied significantly across different studies, ranging from 0.10 to 0.99. These variations are potentially caused by differences in sample sizes, input features, and the structure of the original data. In this study, the XGBoost-based model achieved R^2 values of 0.86 for wind velocity predictions and 0.75 for air temperature predictions on test data, both of which fall within the investigated R^2 range. Additionally, the lower R^2 values observed in the validation data in this study are considerably higher than the lower R^2 values in the investigated models, while the higher R^2 values for the entire data approach the upper values of R^2 in investigated models. Therefore, the optimal XGBoost-based

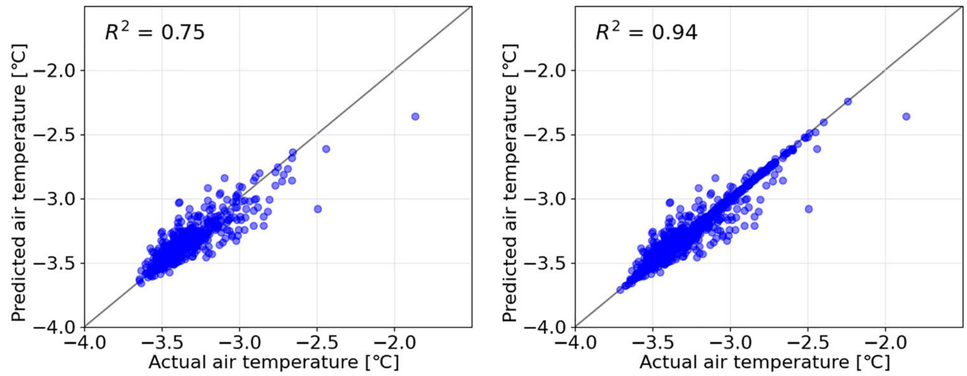


Figure 12. The actual versus the predicted values of the XGBoost-based air temperature model. (a) On test data (b) On entire data.

Table 8. The comparison of R^2 for the optimal models in this study and the similar models in previous studies.

Studies	Year	Target variables	Algorithms	R^2
This study	2025	**Wind velocity**	**XGBoost**	**0.81(V), 0.86(T), 0.97(E)**
		Air temperature	**XGBoost**	**0.70(V), 0.75(T), 0.94(E)**
Ma et al. (2025)	2025	MRT	XGBoost/Simplified energy balance model/ Transformer model	0.58/0.76/0.99
Ma et al. (2025)	2025	Wind velocity	XGBoost/Transformer model	0.55/0.81
Wu, Zhan, and Quan (2021b)	2021	Wind velocity	Response surface	0.84–0.99
Wu, Zhan, and Quan (2021a)	2021	Wind velocity	Gaussian process/Response surface	approximately 0.10–0.90
Du, Mak, and Li (2019)	2019	MVR	Linear model/Quadratic model	0.82/0.98
Du, Mak, and Li (2019)	2019	PET	Linear model/Quadratic model	0.90/0.99

Note: V, T, and E represent Validation, Test, and Entire.

models provided satisfactory performance for efficiently predicting pedestrian level wind velocity and air temperature in this case study. Although these two models may face challenges in predicting wind velocity and air temperature for other cases with significantly different urban morphologies or inflow profiles, the modelling framework developed in this study can be directly utilized for further model development. Furthermore, the two optimal prediction models serve as valuable tools for further analyzing the effect of urban morphology on pedestrian level wind velocity and air temperature.

3.3. SHAP-based effect analysis results

The study utilized SHAP summary plots and dependence plots to analyze the effect of urban morphology on wind velocity and air temperature. The summary plot identifies key design variables that significantly influence wind velocity and air temperature, while the dependence plot captures how the impact of design variables on wind velocity and air temperature interacts with other design variables. Figure 13 presents the SHAP summary plot for wind velocity. The x-axis denotes the SHAP value, quantifying the contribution of the design variable to wind velocity. Positive SHAP values indicate an increase in the target variable, negative values indicate a decrease, and higher absolute SHAP values reflect a greater influence on model output. The y-axis ranks design variables in descending order of feature importance. Each dot in the chart represents a sidewalk sample, where red indicates higher variable values and blue indicates lower values. Figure 14(a and b) display single dependence plots for wind velocity, illustrating how variations in design variable values in the x-axis affect the target variable based on SHAP values in the y-axis. Additionally, Figure 14(c and d) explore how the contribution of one design variable on the target variable interacts with the second design variable. Similar to the single dependence plot, the x-axis and y-axis represent the design variable values and its SHAP values, while the dot gradient colour indicates value changes in the second design variable. The dispersion of red and blue dots highlights the extent to which the effect of design variables on target variables interacts with the second variable.

3.3.1. Analysis of wind velocity

As shown in Figure 13, the summary plot for wind velocity indicates that the horizontal distance, vertical distance, and orientation angle are the three most influential design variables affecting pedestrian level wind velocity. These are followed by the distance between buildings, building height, and building width. Among them, horizontal distance is the most critical design variable, with

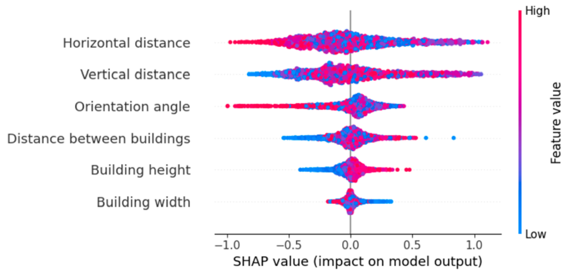


Figure 13. SHAP summary plot for wind velocity.

higher horizontal distances generally leading to lower wind velocity. Vertical distance ranks second in feature importance, with lower vertical distances typically associated with reduced wind velocity. The orientation angle also plays a significant role, as larger orientation angles tend to result in lower wind velocity. In contrast, the other three design variables have a comparatively smaller impact on wind velocity. Specifically, a shorter distance between buildings, lower building height, and greater building width contribute to reducing wind velocity. In summary, the relative position of the sidewalk within the building array centre and the orientation angle exerts a more pronounced influence on pedestrian level wind velocity compared to street width, building height, and building width.

This section discusses the dependence plot, which illustrates the meaningful distribution and dispersion of wind velocity. Figure 14(a and b) present the single dependence plots for orientation angle and building height. The influence of orientation angle on wind velocity exhibits fluctuations, with most angles between 30° and 55° increasing wind velocity, whereas angles exceeding approximately 80° or falling below 10° generally reduce the wind velocity. Building height also plays a crucial role, with 60 m serving as a threshold to distinguish the positive or negative impact of design variables on wind velocity. Buildings taller than 60 m tend to increase wind velocity while those shorter than 60 m typically decrease it. Figure 14(c) highlights how the impact of distance between buildings on wind velocity interacts with the horizontal distance of sidewalks. When the street width is less than approximately 60 m, it more effectively reduces wind velocity for sidewalks with greater horizontal distances than those with shorter ones. Conversely, for streets wider than 60 m, wind velocity tends to increase more significantly for sidewalks with greater horizontal distances than for those with shorter ones. In contrast, Figure 14(d) demonstrates that narrower streets contribute to a greater increase in wind velocity for sidewalks with higher vertical distances, while wider streets tend to decrease their wind velocity more significantly.

3.3.2. Analysis of air temperature

Figure 15 presents the summary plot for air temperature, highlighting horizontal and vertical distance are the two primary design variables influencing pedestrian level air temperature. This finding aligns with the wind velocity analysis. The impact of horizontal and vertical distances on air temperature exhibits a clear directional trend, with greater horizontal distances corresponding to higher air temperatures and lower vertical distances also contributing to increased air temperatures. Given that the inlet wind direction is 0° (North) and the building array rotates from east to north (Figure 2(c)), sidewalks with greater horizontal distances and lower vertical distances receive less shade from surrounding buildings and experience higher solar exposure, ultimately resulting in increased air temperatures. The distance between buildings emerges as the third most influential design variable, with smaller distances generally resulting in higher air temperatures. In contrast, building height, orientation angle, and building width have a relatively minor effect on air temperatures, with shorter buildings, and smaller building widths generally leading to higher air temperatures.

Figure 16 presents the dependence plot for air temperature, illustrating that greater horizontal distances and lower vertical distances can lead to a maximum increase in air temperature of over 0.4°C. Figure 16(c) exhibits a trend similar to that of wind velocity, indicating that most orientation angles between 30° and 55° contribute to an increase in air temperature. Figure 16(d–e) further reveal that a building height of approximately 60 m serves as a threshold in determining its impact on air temperature, with its influence significantly interacting with both horizontal and vertical distances. As shown in Figure 16(d), buildings shorter than 60 m tend to increase the air temperature of sidewalks with lower horizontal distances more than those with higher horizontal distances. Conversely, buildings taller than 60 m reduce the air temperature of sidewalks with lower horizontal distances more significantly than those with higher horizontal distances. Similarly, Figure 16(e) demonstrates that shorter buildings contribute to a greater increase in air temperature for sidewalks with higher vertical distances, whereas taller buildings more effectively decrease their air temperature. Figure 16(f) highlights how the effect of orientation angle on air temperatures interacts with the vertical distance. When the orientation angle is below 30°, the increase in air temperature is more pronounced for sidewalks with lower vertical distances than those with higher vertical distances. In contrast, when the orientation angle exceeds 55°, the reduction in air temperature is more substantial for sidewalks with lower vertical distances compared to those with higher vertical distances.

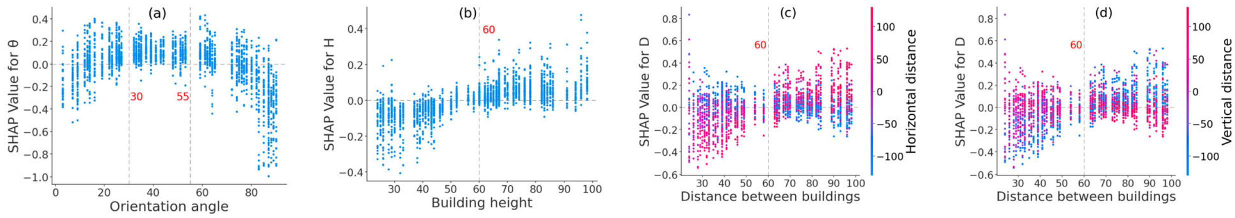


Figure 14. SHAP dependence plot for wind velocity.

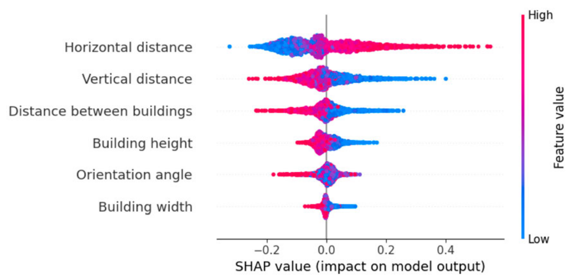


Figure 15. SHAP summary plot for air temperature.

3.4. Policy implication

To enhance pedestrian level wind and thermal comfort during winter in Beijing, it is essential to reduce wind velocity while increasing air temperature. This study employed the SHAP heatmap plot to identify combinations of key design variables that are likely to achieve these objectives. Figure 17 presents the SHAP heatmap plot for wind velocity and air temperature, respectively. These heatmaps visualize the SHAP values in a matrix format, where the x-axis indicates samples of urban morphology, the y-axis denotes design variables, and the colour-encoded SHAP values illustrate the influence of each design variable on wind velocity and air temperature. The right-side bars show the importance of design variables. Above the matrix, the model output is displayed as a descending $f(x)$ curve, providing an overview of the predicted outcomes.

Figure 17(a) displays the SHAP heatmap plot for wind velocity. By analyzing it alongside the corresponding SHAP summary plot, it becomes evident that a combination of a larger horizontal distance, smaller vertical distance, larger orientation angle, narrower street width, lower building height, and greater building width can contribute to reducing pedestrian level wind velocity. Based on this analysis, several design strategies were recommended to potentially mitigate pedestrian level wind velocity during winter in Beijing's urban environment. First, orienting buildings at larger angles to the prevailing wind direction can disrupt direct wind pathways, thereby reducing wind velocity along pedestrian corridors. Second, in specific urban morphologies, narrowing streets may help break wind pathways and reduce pedestrian level wind velocity. Third, reducing building heights can alleviate the downwash effect, thereby contributing to lower pedestrian level wind velocity. Lastly, in specific urban morphologies, increasing building width can help block and diffuse wind flow, further minimizing wind velocity. Integrating these strategies into urban planning and building layout design at an early stage can contribute to creating a more comfortable wind environment for pedestrians.

Figure 17(b) presents the SHAP heatmap plot for air temperature. Similarly, the analysis reveals that a combination of larger horizontal distance, smaller vertical distance, narrower street width, and lower building height and width can contribute to increasing pedestrian level air temperature. Based on this analysis, the following design strategies were recommended to potentially enhance pedestrian level air temperature. First, while wider streets can increase sunlight exposure and raise air temperatures, narrowing streets in certain configurations may reduce wind flow, create sheltered microclimates that retain heat, and consequently increase air temperature. Therefore, optimizing street width requires a careful balance between maximizing sunlight exposure and minimizing wind effects to effectively improve pedestrian level air temperature. Second, reducing building heights and widths decreases shading and enhances solar radiation exposure at the pedestrian level, thereby contributing to higher air temperatures.

4. Conclusion and future work

The wind and thermal environments at the pedestrian level are crucial factors influencing outdoor human comfort. Enhancing wind and thermal environments through urban morphology design has garnered significant attention. However, most existing studies relied on CFD simulations to evaluate pedestrian level wind and thermal conditions, which were both time-intensive and computationally demanding. Additionally, existing studies primarily assessed the impact of urban morphology on wind velocity and air temperature based on predesigned project scenarios, often providing only global coefficients, thereby lacking profound interpretability.

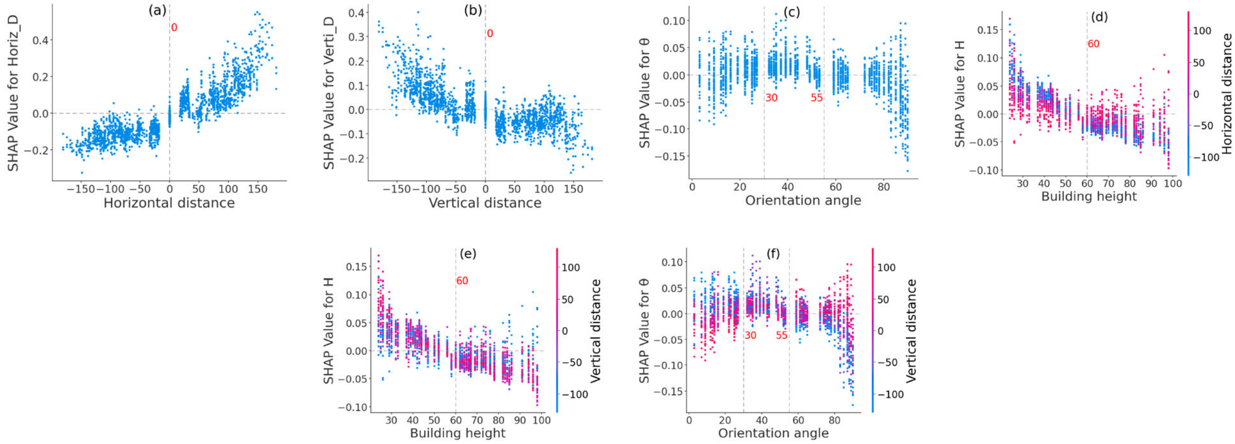


Figure 16. SHAP dependence plot for air temperature.

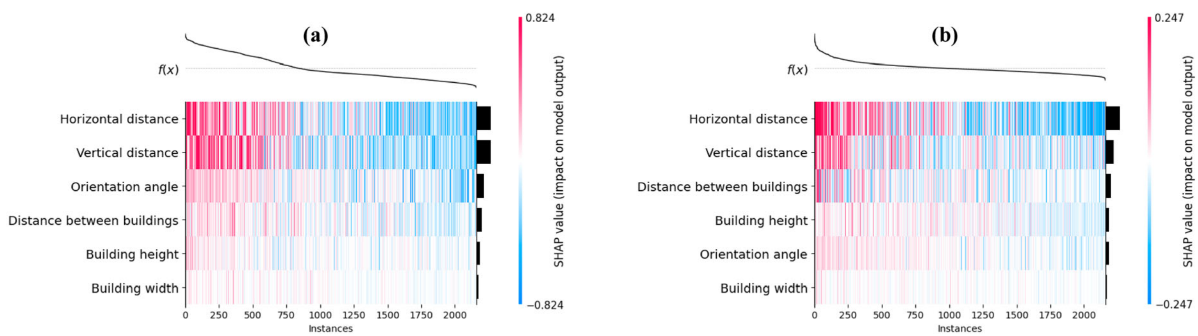


Figure 17. SHAP heatmap plot for (a) wind velocity and (b) air temperature.

To address these limitations, this study leveraged CFD simulation data in combination with two machine learning algorithms to develop models for pedestrian level wind velocity and air temperature predictions. The results indicated that XGBoost-based models outperform LightGBM-based models, achieving R^2 of 0.86 and RMSE of 0.28 m/s for wind velocity, and R^2 of 0.75 and RMSE of 0.11°C for air temperature. SHAP was used for the optimal XGBoost-based models to further analyze the influence of urban morphology on wind velocity and air temperature. The findings revealed that, compared to conventional urban morphology design variables, such as street width, building height, building width, and orientation angle, the relative position of sidewalks within the building array centre plays a more significant role in shaping pedestrian level wind and thermal environments. Additionally, the study investigated the interactions between the effect of one design variable on pedestrian level wind velocity and air temperature and another design variable. Based on the SHAP effect analysis, the study proposed design strategies for the early stage of urban planning and building layout design.

The novelty of this study lies in the two prediction models capable of estimating pedestrian level wind velocity and air temperature within seconds, significantly reducing the reliance on time-intensive simulations. These models serve as rapid and reliable tools for assessing pedestrian level wind and thermal environments in urban areas with morphological characteristics similar to those examined in this study. Furthermore, the SHAP-based effect analysis provided valuable insights into the influence of urban morphology on pedestrian level wind velocity and air temperature, offering practical policy implications for urban planning and building layout design. By effectively applying these findings to real urban environments in Beijing, this research can contribute to creating a more comfortable pedestrian wind and thermal environment while enhancing the efficiency of urban design decision-making.

The study recognizes limitations to be addressed in future research. While two prediction models established in this study effectively estimate wind velocity and air temperature around a 3 × 3 building array, this represents an idealized geometry. Their predictive capability in more complex and diverse urban morphologies remains uncertain, posing challenges in generalizing these models to real-world urban environments. Future research should focus on enhancing the adaptability and practicality of these prediction models for broader urban applications. Additionally, the recommended strategies for each design variable exhibited both similarities and differences in improving pedestrian level wind and thermal environments during winter in Beijing. To further refine these strategies, future studies could integrate wind velocity and air temperature into a unified thermal comfort criterion, offering a more comprehensive approach to optimizing pedestrian level thermal comfort.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability

Data will be made available on request.

References

Architectural Institute of Japan. 2020. *Guidebook for CFD Predictions of Urban Wind Environment*. Tokyo: Architectural Institute of Japan. https://www.aij.or.jp/jpn/publish/cfdguide/index_e.htm.

Celik, Ismail, U. Ghia, P. J. Roache, C. J. Freitas, H. Coloman, and Peter E. Raad. 2008. "Procedure for Estimation and Reporting of Uncertainty Due to Discretization in CFD Applications." *Journal of Fluids Engineering* 130 (7): 078001. https://doi.org/10.1115/1.2960953.

Chen, G., L. Rong, and G. Zhang. 2021. "Unsteady-state CFD Simulations on the Impacts of Urban Geometry on Outdoor Thermal Comfort Within Idealized Building Arrays." *Sustainable Cities and Society* 74:103187. https://doi.org/10.1016/j.scs.2021.103187.

Clemente, A. V., K. E. T. Giljarhus, L. Oggiano, and M. Ruocco. 2024. "Rapid Pedestrian-Level Wind Field Prediction for Early-stage Design Using Pareto-optimized Convolutional Neural Networks." *Computer-Aided Civil and Infrastructure Engineering* 39 (18): 2826–2839. https://doi.org/10.1111/mice.13221.

Cui, X., M. Lee, C. Koo, and T. Hong. 2024. "Energy Consumption Prediction and Household Feature Analysis for Different Residential Building Types Using Machine Learning and SHAP: Toward Energy-efficient Buildings." *Energy and Buildings* 309:113997. https://doi.org/10.1016/j.enbuild.2024.113997.

Dong, W., D. Dai, M. Liu, Y. Wang, S. Li, and P. Shen. 2025. "Combined Effects of the Visual-thermal Environment on Restorative Benefits in Hot Outdoor Public Spaces: A Case Study in Shenzhen, China." *Building and Environment* 272:112690. https://doi.org/10.1016/j.buildenv.2025.112690.

Du, Y., C. M. Mak, K. Kwok, K.-T. Tse, T. Lee, Z. Ai, J. Liu, and J. Niu. 2017. "New Criteria for Assessing Low Wind Environment at Pedestrian Level in Hong Kong." *Building and Environment* 123:23–36. https://doi.org/10.1016/j.buildenv.2017.06.036.

Du, Y., C. M. Mak, and Y. Li. 2019. "A Multi-stage Optimization of Pedestrian Level Wind Environment and Thermal Comfort with Lift-up Design in Ideal Urban Canyons." *Sustainable Cities and Society* 46:101424. https://doi.org/10.1016/j.scs.2019.101424.

Giljarhus, K. E. T., and T.-O. Hågbo. 2024. "Simulation-Based Data-driven Wind Engineering—Analyzing the Influence of Building Proximity and Skyways on Pedestrian Comfort." In *Advances in Computational Mechanics and Applications*, edited by D. Pavlou, H. Adeli, J. A. F. O. Correia, N. Fantuzzi, G. C. Georgiou, K. E. Giljarhus, and Y. Sha, 241–253. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-49791-9_17.

Gromke, C., and B. Blocken. 2015. "Influence of Avenue-trees on Air Quality at the Urban Neighborhood Scale. Part I: Quality Assurance Studies and Turbulent Schmidt Number Analysis for RANS CFD Simulations." *Environmental Pollution* 196:214–223. https://doi.org/10.1016/j.envpol.2014.10.016.

Jeong, J., J. Jeong, M. Lee, J. Lee, and S. Chang. 2022. "Data-driven Approach to Develop Prediction Model for Outdoor Thermal Comfort Using Optimized Tree-type Algorithms." *Building and Environment* 226:109663. https://doi.org/10.1016/j.buildenv.2022.109663.

Kastner, P., and T. Dogan. 2023. "A GAN-based Surrogate Model for Instantaneous Urban Wind Flow Prediction." *Building and Environment* 242:110384. https://doi.org/10.1016/j.buildenv.2023.110384.

Lan, H., H. (Cynthia) Hou, and Z. Gou. 2023. "A Machine Learning Led Investigation to Understand Individual Difference and the Human-environment Interactive Effect on Classroom Thermal Comfort." *Building and Environment* 236:110259. https://doi.org/10.1016/j.buildenv.2023.110259.

Li, Z., H. Zhang, C.-Y. Wen, A.-S. Yang, and Y.-H. Juan. 2020. "Effects of Frontal Area Density on Outdoor Thermal Comfort and Air Quality." *Building and Environment* 180:107028. https://doi.org/10.1016/j.buildenv.2020.107028.

Liu, J., and J. Niu. 2019. "Delayed Detached Eddy Simulation of Pedestrian-level Wind Around a Building Array – The Potential to Save Computing Resources." *Building and Environment* 152:28–38. https://doi.org/10.1016/j.buildenv.2019.02.011.

Liu, J., J. Niu, Y. Du, C. M. Mak, and Y. Zhang. 2019a. "LES for Pedestrian Level Wind Around an Idealized Building Array—Assessment of Sensitivity to Influencing Parameters." *Sustainable Cities and Society* 44:406–415. https://doi.org/10.1016/j.scs.2018.10.034.

Liu, X., H. Tang, Y. Ding, and D. Yan. 2022. "Investigating the Performance of Machine Learning Models Combined with Different Feature Selection Methods to Estimate the Energy Consumption of Buildings." *Energy Building* 273:112408. https://doi.org/10.1016/j.enbuild.2022.112408.

Liu, C., S. Wang, S. Cuomo, and G. Mei. 2021. "Data Analysis and Mining of Traffic Features Based on Taxi GPS Trajectories: A Case Study in Beijing." *Concurrency and Computation* 33 (3): e5332. https://doi.org/10.1002/cpe.5332.

Liu, J., X. Zhang, J. Niu, and K. T. Tse. 2019b. "Pedestrian-level Wind and Gust Around Buildings with a 'Lift-up' Design: Assessment of Influence from Surrounding Buildings by Adopting LES." *Building Simulation* 12 (6): 1107–1118. https://doi.org/10.1007/s12273-019-0541-5.

Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee. 2020. "From Local Explanations to Global Understanding with Explainable AI for Trees." *Nature Machine Intelligence* 2:56–67. https://doi.org/10.1038/s42256-019-0138-9.

Lundberg, S. M., and S.-I. Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In *Advances in Neural Information*

Processing Systems, edited by U. Von Luxburg, I. Guyon, S. Bengio, H. Wallach, and R. Fergus, 4765–4774. New York: Curran Associates Inc. https://doi.org/10.48550/arXiv.1705.07874.

Ma, X., T. Zeng, M. Zhang, P. Zeng, B. Lin, and S. Lu. 2025. "Street Microclimate Prediction Based on Transformer Model and Street View Image in High-density Urban Areas." *Building and Environment* 269:112490. https://doi.org/10.1016/j.buildenv.2024.112490.

Mi, J.-X., A.-D. Li, and L.-F. Zhou. 2020. "Review Study of Interpretation Methods for Future Interpretable Machine Learning." *IEEE Access* 8:191969–191985. https://doi.org/10.1109/ACCESS.2020.3032756.

Ministry of Housing and Urban-Rural Development of the People's Republic of China. 2018. *Standard for Green Performance Calculation of Civil Buildings JGJ/T449-2018*. Beijing: China Architecture & Building Press.

Ministry of Public Security of the People's Republic of China. 2018. *Code for Fire Protection Design of Buildings GB50016-2014*. Beijing: China Planning Press.

Mochida, A., and I. Y. F. Lun. 2008. "Prediction of Wind Environment and Thermal Comfort at Pedestrian Level in Urban Area." *Journal of Wind Engineering and Industrial Aerodynamics* 96 (10–11): 1498–1527. https://doi.org/10.1016/j.jweia.2008.02.033.

Mochida, A., Y. Tominaga, Y. Ishida, T. Ishihara, K. Uehara, H. Kataoka, T. Kurabuchi, et al. 2016. *AIJ Benchmarks for Validation of CFD Simulations Applied to Pedestrian Wind Environment around Buildings*. Tokyo: Architectural Institute of Japan.

Mokhtar, S., M. Beveridge, Y. Cao, and I. Drori. 2021. "Pedestrian Wind Factor Estimation in Complex Urban Environments, in: Asian Conference on Machine Learning." *Proceedings of Machine Learning Research* 157: 486–501.

Moon, J., S. Rho, and S. W. Baik. 2022. "Toward Explainable Electrical Load Forecasting of Buildings: A Comparative Study of Tree-based Ensemble Methods with Shapley Values." *Sustainable Energy Technologies and Assessments* 54:102888. https://doi.org/10.1016/j.seta.2022.102888.

Roache, P. J. 1997. "Quantification of Uncertainty in Computational Fluid Dynamics." *Annual Review of Fluid Mechanics* 29 (1): 123–160. https://doi.org/10.1146/annurev.fluid.29.1.123.

Shapley, L. S. 1953. "A Value for N-person Games." *Contributions to the Theory of Games*:307–317.

Shen, P., and Z. Wang. 2020. "How Neighborhood Form Influences Building Energy Use in Winter Design Condition: Case Study of Chicago Using CFD Coupled Simulation." *Journal of Cleaner Production* 261:121094. https://doi.org/10.1016/j.jclepro.2020.121094.

Shen, P., M. Wang, H. Ma, and N. Ma. 2024. "On the Two-way Interactions of Urban Thermal Environment and Air Pollution: A Review of Synergies for Identifying Climate-Resilient Mitigation Strategies." *Building Simulation* 18: 259–279. https://doi.org/10.1007/s12273-024-1210-x.

Sun, C., W. Lian, L. Liu, Q. Dong, and Y. Han. 2022. "The Impact of Street Geometry on Outdoor Thermal Comfort within Three Different Urban Forms in Severe Cold Region of China." *Building and Environment* 222:109342. https://doi.org/10.1016/j.buildenv.2022.109342.

Tominaga, Y., A. Mochida, R. Yoshie, H. Kataoka, T. Nozu, M. Yoshikawa, and T. Shirasawa. 2008. "AIJ Guidelines for Practical Applications of CFD to Pedestrian Wind Environment around Buildings." *Journal of Wind Engineering and Industrial Aerodynamics* 96 (10–11): 1749–1761. https://doi.org/10.1016/j.jweia.2008.02.058.

Tong, Z., Y. Chen, and A. Malkawi. 2016. "Defining the Influence Region in Neighborhood-Scale CFD Simulations for Natural Ventilation Design." *Applied Energy* 182:625–633. https://doi.org/10.1016/j.apenergy.2016.08.098.

Wang, M., H. Ma, X. Zheng, C. Han, and P. Shen. 2025. "Two-way Coupled Numerical Simulation between Outdoor Thermal Environment and PM2.5 in Urban Blocks." *Building and Environment* 275:112821. https://doi.org/10.1016/j.buildenv.2025.112821.

Weerasuriya, A. U., X. Zhang, B. Lu, K. T. Tse, and C.-H. Liu. 2020. "Optimizing Lift-up Design to Maximize Pedestrian Wind and Thermal Comfort in 'Hot-calm' and 'Cold-windy' Climates." *Sustainable Cities and Society* 58:102146. https://doi.org/10.1016/j.scs.2020.102146.

Wu, Y., Q. Zhan, and S. J. Quan. 2021a. "Improving Local Pedestrian-level Wind Environment Based on Probabilistic Assessment Using Gaussian Process Regression." *Building and Environment* 205:108172. https://doi.org/10.1016/j.buildenv.2021.108172.

Wu, Y., Q. Zhan, and S. J. Quan. 2021b. "A Robust Metamodel-based Optimization Design Method for Improving Pedestrian Wind Comfort in an Infill Development Project." *Sustainable Cities and Society* 72:103018. https://doi.org/10.1016/j.scs.2021.103018.

Yang, A.-S., Y.-H. Juan, C.-Y. Wen, and C.-J. Chang. 2017. "Numerical Simulation of Cooling Effect of Vegetation Enhancement in a Subtropical Urban Park." *Applied Energy* 192:178–200. https://doi.org/10.1016/j.apenergy.2017.01.079.

Yang, S., L. (Leon) Wang, T. Stathopoulos, and A. M. Marey. 2023. "Urban Microclimate and its Impact on Built Environment – A Review." *Building and Environment* 238:110334. https://doi.org/10.1016/j.buildenv.2023.110334.

Yang, Y., Y. Yuan, Z. Han, and G. Liu. 2022. "Interpretability Analysis for Thermal Sensation Machine Learning Models: An Exploration Based on the SHAP Approach." *Indoor Air* 32: e12984. https://doi.org/10.1111/ina.12984.

Zhang, J., P. Cui, and H. Song. 2020. "Impact of Urban Morphology on Outdoor air Temperature and Microclimate Optimization Strategy Base on Pareto Optimality in Northeast China." *Building and Environment* 180:107035. https://doi.org/10.1016/j.buildenv.2020.107035.

Zhang, J., Z. Li, Y. Wei, and D. Hu. 2022. "The Impact of the Building Morphology on Microclimate and Thermal Comfort-a Case Study in Beijing." *Building and Environment* 223:109469. https://doi.org/10.1016/j.buildenv.2022.109469.

Zhang, Y., B. K. Teoh, M. Wu, J. Chen, and L. Zhang. 2023. "Data-driven Estimation of Building Energy Consumption and GHG Emissions Using Explainable Artificial Intelligence." *Energy* 262:125468. https://doi.org/10.1016/j.energy.2022.125468.

Zheng, X., L. Chen, and J. Yang. 2023. "Simulation Framework for Early Design Guidance of Urban Streets to Improve Outdoor Thermal Comfort and Building Energy Efficiency in Summer." *Building and Environment* 228:109815. https://doi.org/10.1016/j.buildenv.2022.109815.