SimBuild
2024

Eleventh National Conference of IBPSA-USA
Denver, Colorado
May 21 – 23, 2024

Effects of Urban Morphology on Pedestrian Level Wind Environment and Air Temperature: Using Simulation and Explainable Machine Learning

Xue CUI[1], Yu LI[23], and Pengyuan SHEN[4]
[1]Department of Building and Real Estate, Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong, China
[2]School of Architecture, Harbin Institute of Technology, Harbin, China
[3]Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, Harbin, China
[4]School of Architecture, Harbin Institute of Technology, Shenzhen, China

Abstract

Urban morphology plays a crucial role in affecting pedestrian level wind velocity and air temperature. Most existing studies utilized the CFD method to simulate environmental parameters at the pedestrian level, while using statistical analysis to assess the impact of urban morphology. Simulation has high requirements in computational time, input parameters, and modeler expertise. Statistical analysis results highly rely on the pre-designed scenarios while lacking a profound interpretation. The study addressed these gaps by using CFD simulation data to establish two separate machine learning models for quickly predicting pedestrian level wind velocity and air temperature. Results showed that XGBoost-based models achieved better performance than LightGBM-based models. Then, explainable SHAP was applied to the optimal XGBoost-based models to investigate the effect of urban morphology on pedestrian level wind velocity and air temperature. SHAP results revealed that the location of sidewalks from the building array center has a more significant impact on wind velocity and air temperature than the street width, building height and width, and orientation angle. In addition, the study explored the effect of interactions between urban morphology-related design variables on pedestrian level wind velocity and air temperature.

Keywords

Wind Velocity, Air Temperature, Urban Morphology, Thermal Comfort, Computational Fluid Dynamics, Machine Learning, SHapley Additive exPlanations

Introduction

In urban outdoor spaces, the wind environment and air temperature at the pedestrian level significantly affect human comfort. As one of the essential factors, urban morphology plays a crucial role in affecting pedestrian level wind velocity and air temperature (Sun et al. 2022). Scholars have conducted a series of research in this field (Du, Mak, and Li 2019; Wu, Zhan, and Quan 2021a; Weerasuriya et al. 2020; Zhang, Cui, and Song 2020), where predicting wind velocity and air temperature and identifying the effects of urban morphology on them are two research priorities.

Existing studies mainly employed experimental and numerical approaches to predict wind velocity and air temperature (Yang et al. 2023). The experimental approach is based on field measurements and wind tunnel tests, while the numerical approach entails computational fluid dynamics(CFD) and data-driven models (Yang et al. 2023). Compared to the experimental approach, the numerical approach shows superiority in modifying experimental parameters, setting numerous scenarios, and conducting comparative analysis (Zheng, Chen, and Yang 2023). As one of the numerical approaches, CFD has been widely used to evaluate pedestrian level wind and thermal environments (Mochida and Lun 2008; Shen and Wang 2020). However, it consumes much computing time, ranging from hours to even days (Liu and Niu 2019; Wu, Zhan, and Quan 2021b). In addition, it requires models with detailed input parameters and modelers with professional knowledge and sufficient expertise (Kastner and Dogan 2023). Meeting these requirements makes predicting wind velocity and air temperature challenging in the early design stage. Novel data-driven models can quickly and accurately predict wind and thermal parameters. It uses machine learning algorithms to develop predictive models to capture the relationship between known input urban morphologies and output wind and thermal parameters. It then applies this predictive model to other known input urban morphologies to predict unknown wind and thermal parameters. Yang et al. (2023) indicated that weather station records, field measurements, wind tunnel tests, and CFD simulations are the four primary data sources for developing machine learning models. The coefficient of determination(R^2) of this kind of model is around 0.8-0.9, indicating high prediction accuracy (Yang et al.

2023). In addition, data-driven models show high efficiency in predictions, with a study showing a reduction in computational time from $\sim 8h$ for CFD to $\sim 4s$ for data-driven models (Kastner and Dogan 2023). Although data-driven models offer high prediction accuracy and computational efficiency, limited studies use this approach to predict pedestrian level wind and thermal conditions (Yang et al. 2023). Therefore, there is a need to develop machine learning models to efficiently predict reliable wind velocity and air temperature at the pedestrian level, and to assess the effect of urban mopholpgy on them.

When exploring the effect of urban morphology on pedestrian level wind environment and air temperature, most studies used comparative analysis to explore the environmental parameter differences under different urban characteristics. For instance, Li et al. (2020) used CFD to simulate the outdoor wind velocity and air temperature of five building array scenarios, analyzing the effect of frontal area density of the building array on pedestrian wind and thermal environment. This kind of study commonly simulated environmental parameters under a variety of scenarios one by one, followed by statistical analysis to elucidate how urban morphology affects 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 mechanism of the urban morphology. For instance, Zhang, Cui, and Song (2020) used R^2 to explore the relationship between building layout and outdoor wind comfort. However, such a global coefficient makes it hard to explain the underlying reasons on the one hand, and the impact on specific scenarios on the other. 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 machine learning models and plays a crucial role in the interpretable analysis field (Mi, Li, and Zhou 2020). Most studies have applied SHAP at the building level (Lan, Hou, and Gou 2023; Yang et al. 2022), while few studies used SHAP at the urban level. Therefore, applying SHAP to machine learning prediction models offers an explainable way to explore the relationship between urban morphology and pedestrian level wind velocity or air temperature.

In summary, existing studies predicting pedestrian level wind velocity and air temperature and investigating the effects of urban morphology on them entailed two research gaps. Firstly, the simulation approach for assessing wind and thermal environments is challenging due to the high requirements of computational time and modeler knowledge and expertise. Secondly, statistical analysis and global coefficient are limited in explaining the effect of urban morphology on wind velocity and air temperature in terms of scenario diversity, time limitation, and interpretability. Based on the 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 model to investigate the effect of urban morphology on pedestrian level wind velocity and air temperature in an explainable way. The study is organized into three main sections to achieve these objectives: (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 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.

Methodology

CFD simulation

CFD simulation based on the Ansys Fluent 2019 R3 was selected to calculate pedestrian level wind velocity and air temperature under different urban morphologies. The obtained dataset is further used to develop machine learning prediction models and SHAP analysis.

Urban geometry modeling

The study modeled a building array with nine cuboids as an ideal urban model (Figure 1), which was based on the CaseC of the Architectural Institute of Japan (AIJ) (Tominaga et al. 2008). As the study focuses on pedestrian level conditions, sidewalks around buildings are considered test areas, as shown in the black-filled areas in Figure 1(b). Since the spatial distributions of wind velocity and air temperature vary considerably, the study set 36 test surfaces in total for all four directions of buildings. This can help monitor the environmental parameter differences in spatial distributions. These sidewalks are designed to be 5m away from building surfaces, 1.5m above ground level, 5m wide, and the same length as the building width.

Based on the AIJ guidelines (Architectural Institute of Japan 2016), the computational domain of the CFD simulation was set based on the building height(H). As shown in Figure 1(a), the distance between the inlet and the first building windward surface is 5H, the distance between side symmetry to building sidewalls is 5H, the distance between the outlet and the last building leeward surface is 15H, the top symmetry is 8H away from the building roof.

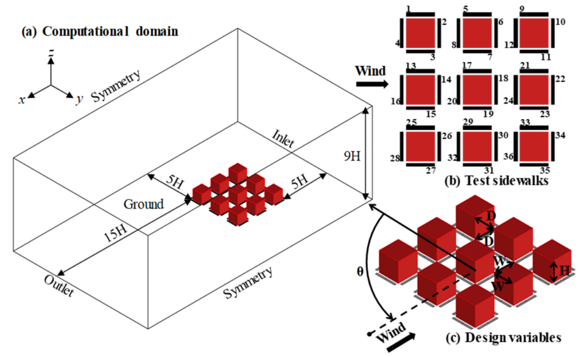


Figure 1 Description of urban geometry modeling

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 design variables have significant effects on pedestrian level wind and thermal environment (Wu, Zhan, and Quan 2021a; Du, Mak, and Li 2019). To generate various urban morphologies, the study set the range and discretization for four design variables (Table 1). Then, the Latin hypercube sampling (LHS) was utilized to generate unique design schemes randomly as it can generate limited samples to better cover the high-dimensional design space (Wu, Zhan, and Quan 2021a). The range of H, W, and D refers to the "Code for fire protection design of buildings GB50016-2014" (Ministry of Public Security of the People's Republic of China 2018) and the existing reference (Du, Mak, and Li 2019). Since design schemes randomly generated by LHS are not ideally symmetric in θ and 90-θ degrees, the study defined the range of θ as 0° to 90°. The mean wind velocity(M_Vel) and air temperature(M_Ta) in each test sidewalk are selected as the target variables.

Table 1 Description of design variables

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

CFD validation

To select the suitable turbulence model and mesh scheme, the study conducted the CFD validation based on the wind tunnel results of AIJ CaseC, which is available on an online AIJ website (Architectural Institute of Japan 2020). CaseC geometry and test points are shown in Figure 2(a). Detailed inflow profiles provided by AIJ were used as boundary conditions for CFD validation (Figure 2(b)-(c)). Wind tunnel results at the height of 0.1H were compared with CFD validation results.

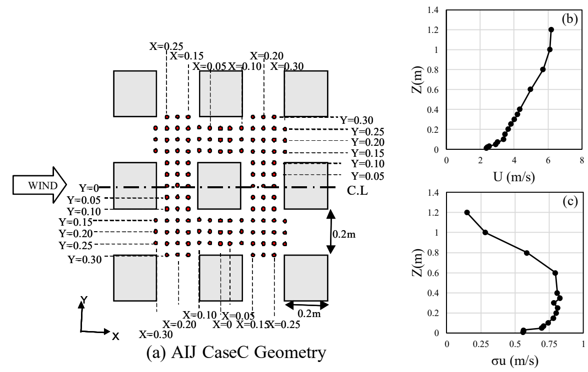


Figure 2 Description of AIJ CaseC

Three mesh systems, including coarse, medium, and fine, were established for the mesh sensitivity test. The growth rate of the three mesh systems was set to 1.2, and all mesh systems had ten inflation layers with the first layer thickness of 0.0011m. The global element sizes for the coarse, medium, and fine systems were 0.054m, 0.036m, and 0.024m, respectively, while the element sizes around their buildings were set to 0.027m, 0.018m, and 0.012m, respectively. In addition, three k-ε models, including standard k-ε, RNG k-ε, and realizable k-ε, were compared to select the most suitable turbulence model. The solution method was based on the coupled and second-order upwind schemes. The study used the Spearman correlation coefficient (Cc) to measure the correlation between the wind tunnel and the CFD validation results with different mesh systems and k-ε models. Table 2 summarizes nine results. As the standard k-ε with the medium mesh outperformed other results, the study selected the standard k-ε model and the medium mesh generation scheme for further CFD simulation.

Table 2 Cc between wind tunnel and CFD validation

k-ε model	Mesh system	Correlation coefficient
Standard k-ε	Coarse	0.83
	Medium	***0.84***
	Fine	0.83
RNG k-ε	Coarse	0.59
	Medium	0.55
	Fine	0.48
Realizable k-ε	Coarse	0.82
	Medium	0.83
	Fine	0.82

CFD simulation

The study applied the typical urban geometry of AIJ CaseC to Beijing's weather conditions as a case study. Beijing was selected as the study area, which is the capital of China and represents the typical city that has experienced intensive urbanization in the past 30 years in China (Zhang et al. 2022). Studying the effects of urban morphology on pedestrian level wind velocity and air temperature is significant in Beijing as it is highly urbanized, with diverse and complex buildings in urban areas (Zhang et al. 2022). The study focused on winter conditions at the pedestrian level as Beijing is located in the cold climate zone of China, with an average outdoor air temperature of -0.7℃ in winter (Yin et al. 2021).

Based on the CFD validation results, the study utilized the standard k-ε model for wind environment simulation, while the energy model and discrete ordinates(DO) model were used to evaluate the radiant heat fluxes between surfaces as it can achieve high accuracy (Li et al. 2020).

For k-ε model, the vertical inflow wind velocity $U(z)$, turbulent kinetic energy $k(z)$ and the turbulence dissipation rate $\varepsilon(z)$ profiles followed the equations (1)-(4) proposed by AIJ (Tominaga et al. 2008):

$$U(z) = U_s(\frac{z}{z_s})^{\alpha} \tag{1}$$

$$I(z) = \frac{\sigma_u(z)}{U(z)} = 0.1(\frac{z}{z_G})^{(-\alpha-0.05)} \tag{2}$$

$$k(z) = \frac{\sigma_u{}^2(z) + \sigma_v{}^2(z) + \sigma_w{}^2(z)}{2} \cong \sigma_u{}^2(z) = (I(z)U(z))^2 \tag{3}$$

$$\varepsilon(z) = C_\mu{}^{1/2}k(z)\frac{U_s}{z_s}\alpha(\frac{z}{z_s})^{(\alpha-1)} \tag{4}$$

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

For energy and DO models, the properties of wall materials are based on the literature (Li et al. 2020). The heat transfer coefficients (h_c) of ground and building walls are calculated by the suggested empirical equation (Yang et al. 2017):

$$h_c = 5.7 + 3.8V_{air} \tag{5}$$

Where V_{air} is the airflow velocity.

Winter in Beijing usually starts in December and ends in February. To determine the simulation time, the study calculated the mean Universal Thermal Climate Index (UTCI) for each hour of the winter daytime based on the local EPW file. Figure 3 shows that the mean UTCI is lowest at 9 am in winter daytime, which is during the morning rush hour in Beijing (Liu et al. 2021). This implies that the thermal comfort of pedestrians is the worst during the peak hour of 9 am. Therefore, 9 am was selected as the simulation time. Based on the local EPW file, the study calculated the mean values of wind velocity, air temperature, and solar radiation-related parameters at 9 am in winter in Beijing, and the mode of wind direction during this period as the simulation inflow parameters. The simulation date was selected based on the typical period in the local EPW file. The power-law exponent and atmospheric boundary layer height were set based on 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.

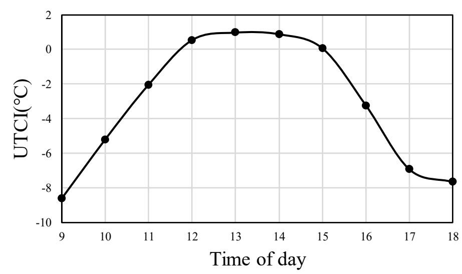


Figure 3 Mean UTCI in winter daytime

Table 3 CFD simulation parameters and settings

Parameters	Settings
Date and time	9 am, 14th December
Inlet	α=0.22, z_s=10m, U_s=2.69m/s, z_G=400m, wind direction=0°, air temperature=-3.81℃
Ground and building wall	No-slip stationary wall
Lateral and top boundary	Symmetry
Outlet	Outflow

The simulation used a 12th Gen Intel(R) Core(TM) i7-12700K-3.60 GHz, 32 GB RAM high-performance computer with parallel processing. To ensure the reliability of CFD simulation in Beijing, the settings of the computational domain, mesh generation scheme, turbulence model, and solution method were the same as the CFD validation. In addition, the study conducted the mesh sensitivity test for the simulation case in Beijing since this case has different inflow profiles, where, notably, the mesh generation method is the same as the validation case. Eight mesh systems, from coarse to fine, are generated in total. The growth rate of each mesh system was set to 1.2, and all mesh systems had ten inflation layers with the first layer thickness of 0.027m.

The global element sizes from mesh 1 to mesh 8 were 39m, 36m, 33m, 30m, 27m, 24m, 21m, and 18m, respectively. In addition, the element sizes around buildings were set to 6.5m, 6m, 5.5m, 5m, 4.5m, 4m, 3.5m, and 3m, respectively. Similarly, the study adopted the Cc to measure the correlation between results generated by the finest mesh system 8 and other mesh systems. Table 4 summarizes the total number of mesh elements, simulation running time, and Cc between each mesh system and mesh system 8. The mesh system results show a higher variation in wind velocity than air temperature. The study finally selected mesh system 5 for simulation as it achieved high Cc for wind velocity and air temperature with the acceptable number of mesh elements. This means that mesh system 5 can simulate reliable results in a shorter time.

Table 4 The number of mesh elements, simulation running time, and Cc between each mesh system and mesh system 8

Mesh	Number of mesh elements	Running time	Cc with mesh system 8	
			Wind velocity	Air temperature
1	574,371	1.40h	0.90	0.98
2	869,638	2.12h	0.90	0.98
3	1,038,649	2.30h	0.90	0.98
4	1,336,391	3.00h	0.92	0.98
5	***1,768,283***	***3.97h***	***0.93***	***0.99***
6	2,278,086	4.45h	0.93	0.99
7	3,223,059	7.07h	0.97	1.00
8(Finest)	4,901,300	22.35h	1.00	1.00

Based on the settings of design variables and target variables discussed earlier, 60 design schemes of urban morphology were randomly generated based on LHS. Figure 4 shows some geometries of the 60 design schemes sampled with LHS. All these 60 design schemes followed the simulation parameters and settings in Table 3 and the mesh system 5. Since each design scheme has 36 sidewalk test surfaces, the study added the horizontal distance (Horiz_D) and vertical distance (Verti_D) between each sidewalk center and the center of the building array to design variables (Figure 5). Meanwhile, the directionality of the horizontal and vertical distances was defined based on the relative position of the sidewalk to the center of the building array. 2,160 (60*36) sidewalk samples were ultimately generated in total. Each sample includes design variable data, horizontal and vertical distance data from the building array center, and corresponding wind velocity and air temperature. Using this dataset for analysis can capture the effect of urban morphology on target variables while identifying spatial distribution variation of target variables.

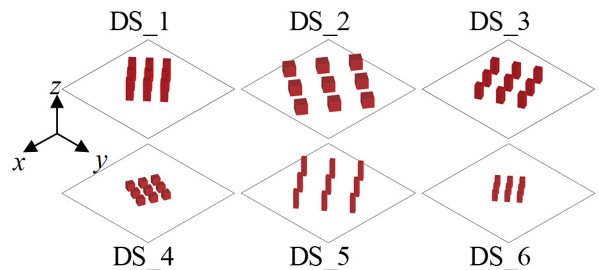


Figure 4 Geometries of some design schemes sampled with LHS

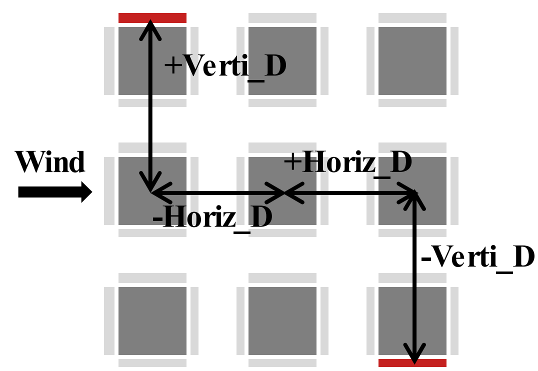


Figure 5 Horizontal and vertical distance

Prediction model development

The study selected tree-based algorithms to develop the machine learning model for predicting wind velocity and air temperature at the pedestrian level. eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) were selected as they achieved superior performance in previous studies (Liu et al. 2022; Moon, Rho, and Baik 2022). The study split the entire dataset into 80% training and 20% test sets. Ten-fold cross-validation was used on the training set to find the optimal model hyperparameter. The model was then evaluated on the test set to explore the prediction ability of this model on the unknown data. As the study used SHAP to analyze the effect of urban morphology on pedestrian level wind velocity and air temperature, the entire dataset was more comprehensive in capturing their relationship than the training or test set. Consequently, the model performance was further evaluated on the entire dataset.

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 a model (Zhang et al. 2023), while RMSE calculates the average difference between predicted and actual values (Liu et al. 2022). Higher R^2 and lower RMSE indicate better model performance. The calculation of two metrics is shown in equations (6)-(7):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{Prediction,i} - y_{Actual,i})^2}{\sum_{i=1}^{n}(y_{Actual,i} - \bar{y}_{Actual})^2} \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{Prediction,i} - y_{Actual,i})^2} \tag{7}$$

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

Effects analysis using SHAP

Existing machine learning models are often black-box models, leading to difficulties in interpreting model predictions. The classic game-theoretic Shapley value can interpret the contribution of model input on model output based on the marginal contribution of features (Shapley 1953). Equation (8) defines the calculation of Shapley value $\emptyset_i$ (Lundberg and Lee 2017):

$$\emptyset_i = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!\,(|F|-|S|-1)!}{|F|!}\left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)\right] \tag{8}$$

Where F is the set including all features, S is the subset excluding the ith feature, and $f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)$ is the prediction difference, including and excluding the ith feature.

Based on the Shapley value, Lundberg and Lee proposed SHAP in 2017, which inherited all the advantages of Shapley value and improved the computation speed of the Shapley value from the exponential to the polynomial level (Lundberg et al. 2020). Consequently, SHAP can offer local explanations for a specific sample prediction while capturing the global structure (Lundberg et al. 2020), where SHAP value is used to measure the feature contribution. The study employed SHAP to the optimal wind velocity and air temperature prediction models to interpret the effects of urban morphology on wind velocity and air temperature from global and local aspects.

Discussion and result analysis

CFD simulation results

Table 5 summarizes simulation results for 2,160 sidewalk samples. Figure 6 shows the distribution of the simulated mean air temperature and wind velocity for 2,160 sidewalk samples. Compared to the inflow air temperature of -3.81°C and wind velocity of 2.69m/s, urban morphology showed a more significant effect on the wind environment as the median value of simulated wind velocity is 0.98m/s, and the median value of simulated air temperature is -3.35°C.

Table 5 Statistical description of simulation results

	Mean	Standard Deviation	Min	Max
H	58.42	20.33	24	98
W	35.28	9.29	20	50
D	60.83	22.62	24	98
θ	47.70	25.93	3	90
Horiz_D	|68.27|	|47.71|	0	|180.5|
Verti_D	|68.27|	|47.71|	0	|180.5|
M_Ta	-3.32	0.20	-3.71	-1.87
M_Vel	1.24	0.77	0.09	3.30

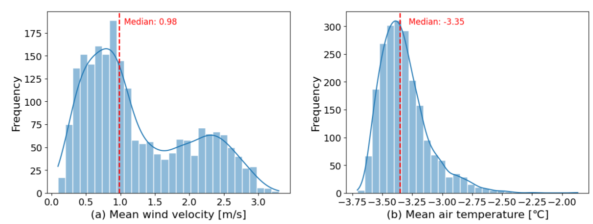


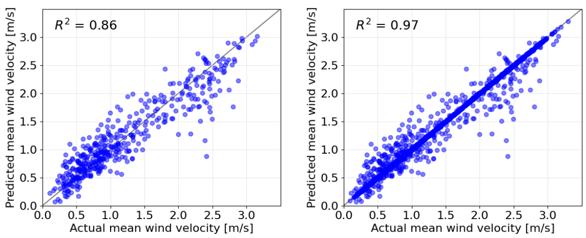
Figure 6 Distribution of target variables

Prediction model results

As mentioned, the model performance was evaluated on the test set to assess the model prediction ability and the entire data to support SHAP analysis. In addition, the study listed the validation results of developed models. Table 6 summarizes the performance of wind velocity prediction models. In summary, two models achieved satisfactory performance, with R^2 around 0.80 on validation data, around 0.85 on test data, and over 0.96 on entire data, while RMSE was below 0.35m/s on validation data, below 0.29m/s on test data, and below 0.16m/s on entire data. Specifically, XGBoost-based wind velocity prediction model performed better on validation, test, and entire data. Therefore, the study selected the XGBoost-based wind velocity model as the optimal pedestrian level wind velocity prediction model. Figure 7 displays the actual mean wind velocity versus the predicted mean wind velocity on the test and entire data, where most samples match the diagonal except for a few points.

Table 6 Performance of wind velocity prediction models

		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



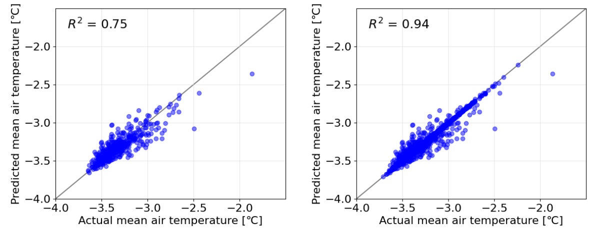
(a) On the test data *(b) On the entire data*

Figure 7 The actual versus the predicted value of XGBoost-based wind velocity model

Table 7 summarizes the performance of air temperature prediction models. The two models also achieved satisfactory results, with the R^2 around 0.70 on validation data, around 0.75 on test data, and over 0.90 on entire data, while RMSE of 0.11℃ on validation and test data, and around 0.05℃ on entire data. Like the wind velocity, the XGBoost-based air temperature model slightly outperformed the LightGBM-based model regarding validation, test, and entire data. Therefore, the study selected the XGBoost-based air temperature model as the optimal pedestrian level air temperature prediction model. Figure 8 shows the actual mean air temperature versus the predicted mean air temperature on the test and entire data, with most of the samples showing high goodness of fit, except for a few points.

Table 7 Performance of air temperature prediction models

		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



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

Figure 8 The actual versus the predicted value of the XGBoost-based air temperature model

Effects analysis results

The study employed the SHAP summary plot and dependence plot to investigate the effect of urban morphology on wind velocity and air temperature. The summary plot identifies key design variables that significantly affect wind velocity and air temperature, while the dependence plot captures the impact of interactions between design variables on wind velocity and air temperature. Figure 9 shows the summary plot for wind velocity. The x-axis uses the SHAP value to quantify the contribution of design variables on wind velocity, with positive SHAP values increasing target variables, negative SHAP values decreasing target variables, and higher absolute SHAP values representing higher contributions to model output. The y-axis lists the design variables based on the descending order of feature importance. Each dot in the chart represents a sidewalk sample, with red representing a higher variable value and blue representing a lower variable value. Figure 10(a)-(b) shows the single dependence plot for wind velocity. It can investigate how the value of design variables in the x-axis affects the target variable based on SHAP value of the y-axis. Figure 10(c)-(d) were used to explore how the contribution of the design variable on the target variable interacts with the second design variable. The x-axis and y-axis explanations are similar to the single dependence plot, while the gradient color of the dots represents the value change of the second design variable on the right side of the chart. The dispersion of red and blue dots indicates the effect of design variables on target variables interacting with the second variable.

Analysis of wind velocity

As shown in Figure 9, the summary plot for wind velocity revealed that the horizontal distance, vertical distance, and orientation angle are the first three essential design variables affecting the pedestrian level wind velocity, followed by the distance between buildings, building height, and width. Horizontal distance is the most critical design variable, with most higher horizontal distances resulting in lower wind velocity. Vertical distance is the second critical design variable, with most lower vertical distances resulting in lower wind velocity. The orientation angle also greatly influences wind velocity, with most larger orientation angles more likely reducing wind velocity. The other three design variables have less impact on wind velocity, with lower distance between buildings, lower building height, and greater building width decreasing wind velocity. In summary, the relative position of the sidewalk to the building array center and orientation angle have a more significant impact on the pedestrian level wind velocity, compared to the street width, building height, and building width.

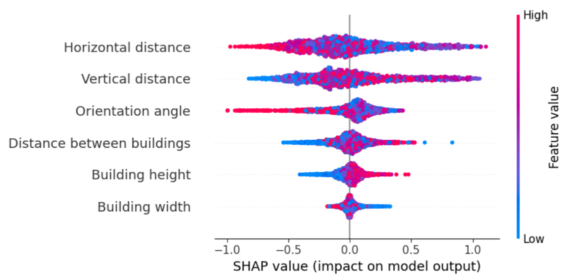


Figure 9 Summary plot for wind velocity

The dependence plot that showed the meaningful distribution and dispersion was discussed in this section. Figure 10(a)-(b) show the single dependence plot of orientation angle and building height. The influence of orientation angle on wind velocity fluctuates with angle. Most angles within 30° to 55° increase wind velocity, while most angles larger than approximately 80° or smaller than 10° reduce wind velocity. A building height of 60m serves as a cutoff point to distinguish the positive or negative impact on wind velocity, with most buildings taller than 60m increasing wind velocity and most buildings shorter than 60m decreasing wind velocity. Figure 10(c) reveals that the impact of distance between buildings on wind velocity interacts with the horizontal distance of sidewalks. Streets that are approximately narrower than 60m can decrease the wind velocity of sidewalks with higher horizontal distances more significantly than lower horizontal ones. Similarly, streets wider than 60m can increase the wind velocity of sidewalks with higher horizontal distances more significantly than lower horizontal ones. In contrast, Figure 10(d) illustrates that narrower streets lead to a higher increase in wind velocity of sidewalks with higher vertical distances, while wider streets more significantly decrease their wind velocity.

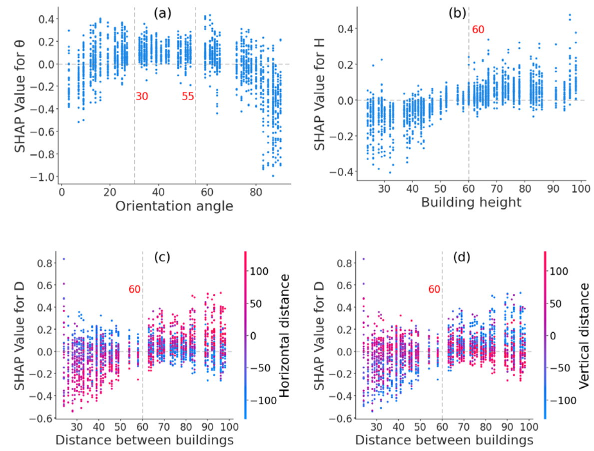


Figure 10 Dependence plot for wind velocity

Analysis of air temperature

Figure 11 shows the summary plot for air temperature. Horizontal and vertical distance are the two key design variables affecting the pedestrian level air temperature, which aligns with the wind velocity analysis. The impact of horizontal and vertical distances on air temperature shows more explicit directionality, with higher horizontal distances leading to higher air temperatures and lower vertical distances leading to higher air temperatures. Given that the inlet wind direction is 0° (North) and the building array rotates from east to north (Figure 1(c)), the sidewalks with higher horizontal distances and lower vertical distances can receive less shade from surrounding buildings and experience greater solar exposure, thus leading to higher air temperature. The distance between buildings is the third key design variable, with higher distances leading to lower air temperatures. Building height, orientation angle, and building width show less impact on air temperatures, with higher building height, most larger orientation angles, and higher building width resulting in lower air temperatures.

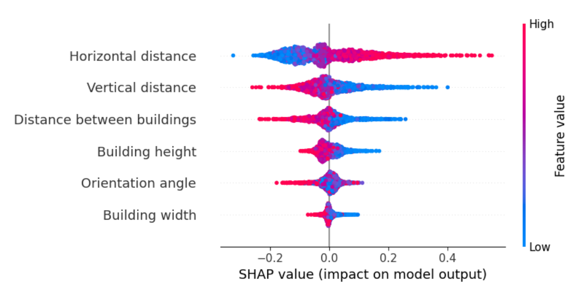


Figure 11 Summary plot for air temperature

Figure 12 shows the dependence plot for air temperature. Higher horizontal and lower vertical distances can result in a maximum increase in air temperatures of over 0.4℃. Figure 12(c) shows a similar effect with the wind velocity, showing that most angles within 30° to 55° can increase air temperature. Figure 12(d)-(e) reveals that a building height of 60m is an approximate cutoff point to distinguish its impact on air temperature. In addition, its impact significantly interacted with the horizontal and vertical distances. As shown in Figure 12(d), buildings approximately shorter than 60m can lead to a higher increase in the air temperature of sidewalks with lower horizontal distances than higher horizontal ones. Similarly, buildings taller than 60m decrease the air temperature of sidewalks with lower horizontal distances more significantly than higher horizontal distances. In contrast, Figure 12(e) reveals that shorter buildings lead to a higher increase in the air temperature of sidewalks with higher vertical distances, while taller buildings more significantly decrease their air temperature. Figure 12(f) reveals that the effect of orientation angle on air temperatures interacts with the vertical distance. An orientation angle smaller than 30° can increase the air temperature of sidewalks with lower vertical distances more significantly than higher vertical distances, while an orientation angle larger than 55° can decrease the air temperature of sidewalks with lower vertical distances more significantly than higher vertical distances.

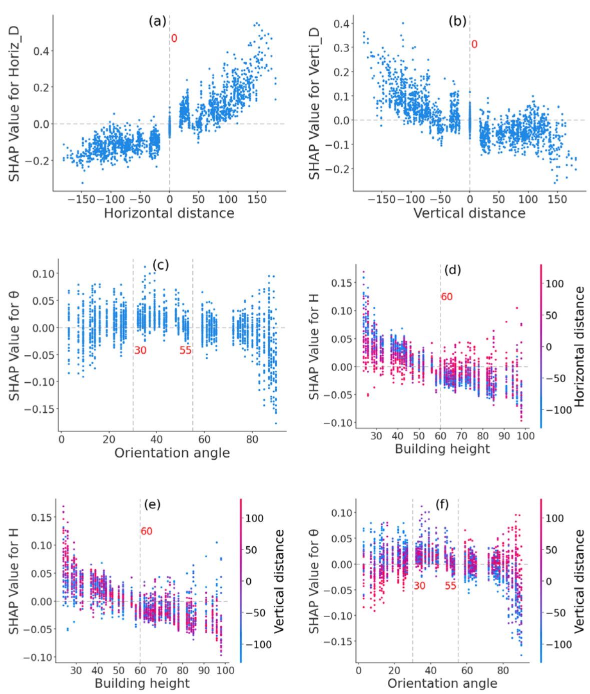


Figure 12 Dependence plot for air temperature

Conclusion

The wind and thermal environments at the pedestrian level play an essential role in human comfort in outdoor spaces. Improving the wind and thermal environments from the urban morphology perspective has received much attention. Most existing studies utilized the CFD method to simulate the pedestrian level wind and thermal parameters, which is time-consuming and has high simulation requirements. In addition, studies highly relied on pre-designed project scenarios to assess the impact of urban morphology on pedestrian level wind and thermal environments. It commonly offered a global coefficient or trend change to interpret the effect, lacking a profound interpretation. The study addressed these gaps by applying the CFD simulation data to two machine learning algorithms to establish models for predicting pedestrian level wind velocity and air temperature. The results showed that XGBoost-based models outperformed LightGBM-based models for predicting wind velocity and air temperature, achieving R^2 of 0.86, RMSE of 0.28m/s for wind velocity, and R^2 of 0.75, RMSE of 0.11°C for air temperature. The study then applied SHAP to the two optimal XGBoost-based prediction models to investigate the impact of urban morphology on pedestrian level wind velocity and air temperature. Compared to urban morphology-related design variables, like street width, building height and width, and orientation angle, the relative position of sidewalks from the building array center plays a more essential role in affecting wind velocity and air temperature. In addition, the study explored the effect of interactions between design variables on pedestrian level wind velocity and air temperature.

This study acknowledges certain limitations that should be addressed in future research. The two prediction models developed in this study can well predict wind velocity and air temperature around a 3×3 building array. However, it is the ideal geometry. The predictive capabilities of these two models remain uncertain in other real-world urban morphologies. Thus, there may be limitations in generalizing the two prediction models to more complex and realistic urban scenarios. Future research should focus on improving the practicality of these prediction models.

References

Architectural Institute of Japan. 2016. AIJ Benchmarks for Validation of CFD Simulations Applied to Pedestrian Wind Environment around Buildings.

Architectural Institute of Japan. 2020. "Guidebook for CFD Predictions of Urban Wind Environment." 2020.

https://www.aij.or.jp/jpn/publish/cfdguide/index_e.htm.

Du, Yaxing, Cheuk Ming Mak, and Yantong 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 (April): 101424. https://doi.org/10.1016/j.scs.2019.101424.

Kastner, Patrick, and Timur Dogan. 2023. "A GAN-Based Surrogate Model for Instantaneous Urban Wind Flow Prediction." Building and Environment 242 (August): 110384. https://doi.org/10.1016/j.buildenv.2023.110384.

Lan, Haifeng, Huiying (Cynthia) Hou, and Zhonghua 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 (May): 110259. https://doi.org/10.1016/j.buildenv.2023.110259.

Li, Zhengtong, Hao Zhang, Chih-Yung Wen, An-Shik Yang, and Yu-Hsuan Juan. 2020. "Effects of Frontal Area Density on Outdoor Thermal Comfort and Air Quality." Building and Environment 180 (August): 107028. https://doi.org/10.1016/j.buildenv.2020.107028.

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

Liu, Jianlin, and Jianlei Niu. 2019. "Delayed Detached Eddy Simulation of Pedestrian-Level Wind around a Building Array – The Potential to Save Computing Resources." Building and Environment 152 (April): 28–38. https://doi.org/10.1016/j.buildenv.2019.02.011.

Liu, Xue, Hao Tang, Yong Ding, and Da Yan. 2022. "Investigating the Performance of Machine Learning Models Combined with Different Feature Selection Methods to Estimate the Energy Consumption of Buildings." Energy and Buildings 273 (October): 112408. https://doi.org/10.1016/j.enbuild.2022.112408.

Lundberg, Scott M., Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. "From Local Explanations to Global Understanding with Explainable AI for Trees." Nature Machine Intelligence 2 (1): 56–67. https://doi.org/10.1038/s42256-019-0138-9.

Lundberg, Scott M, and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In Advances in Neural Information Processing Systems. Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

Mi, Jian-Xun, An-Di Li, and Li-Fang Zhou. 2020. "Review Study of Interpretation Methods for Future Interpretable Machine Learning." IEEE Access 8: 191969–85. 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."

Ministry of Public Security of the People's Republic of China. 2018. "Code for Fire Protection Design of Buildings GB50016-2014."

Mochida, Akashi, and Isaac 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.

Moon, Jihoon, Seungmin Rho, and Sung Wook 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 (December): 102888. https://doi.org/10.1016/j.seta.2022.102888.

Shapley, L S. 1953. "A Value for N-Person Games." Contrib. Theor. Games 2, 307–17.

Shen, Pengyuan, and Zheng 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 (July): 121094. https://doi.org/10.1016/j.jclepro.2020.121094.

Sun, Cheng, Wuyue Lian, Lei Liu, Qi Dong, and Yunsong 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 (August): 109342. https://doi.org/10.1016/j.buildenv.2022.109342.

Tominaga, Yoshihide, Akashi Mochida, Ryuichiro Yoshie, Hiroto Kataoka, Tsuyoshi Nozu, Masaru Yoshikawa, and Taichi 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): 1749–61. https://doi.org/10.1016/j.jweia.2008.02.058.

Weerasuriya, A.U., Xuelin Zhang, Bin Lu, K.T. Tse, and Chun-Ho 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 (July): 102146. https://doi.org/10.1016/j.scs.2020.102146.

Wu, Yihan, Qingming Zhan, and Steven Jige Quan. 2021a. "A Robust Metamodel-Based Optimization Design Method for Improving Pedestrian Wind Comfort in an Infill Development Project." Sustainable Cities and Society 72 (September): 103018. https://doi.org/10.1016/j.scs.2021.103018.

Wu, Yihan, Qingming Zhan, and Steven Jige Quan. 2021b. "Improving Local Pedestrian-Level Wind Environment Based on Probabilistic Assessment Using Gaussian Process Regression." Building and Environment 205 (November): 108172. https://doi.org/10.1016/j.buildenv.2021.108172.

Yang, An-Shik, Yu-Hsuan Juan, Chih-Yung Wen, and Chao-Jui Chang. 2017. "Numerical Simulation of Cooling Effect of Vegetation Enhancement in a Subtropical Urban Park." Applied Energy 192 (April): 178–200. https://doi.org/10.1016/j.apenergy.2017.01.079.

Yang, Senwen, Liangzhu (Leon) Wang, Ted Stathopoulos, and Ahmed Moustafa Marey. 2023. "Urban Microclimate and Its Impact on Built Environment – A Review." Building and Environment 238 (June): 110334. https://doi.org/10.1016/j.buildenv.2023.110334.

Yang, Yuren, Ye Yuan, Zhen Han, and Gang Liu. 2022. "Interpretability Analysis for Thermal Sensation Machine Learning Models: An Exploration Based on the SHAP Approach." Indoor Air 32 (2). https://doi.org/10.1111/ina.12984.

Yin, Peng, Ying Ji, Jingchao Xie, Jiaping Liu, Qixian Hou, and Shanshan Zhao. 2021. "Research on Residential Thermal Environment Differences in Relation to Transition and Heating Periods in Beijing." Building and Environment 197 (June): 107851. https://doi.org/10.1016/j.buildenv.2021.107851.

Zhang, Jiaolong, Zhen Li, Yaxi Wei, and Dan Hu. 2022. "The Impact of the Building Morphology on Microclimate and Thermal Comfort-a Case Study in Beijing." Building and Environment 223 (September): 109469. https://doi.org/10.1016/j.buildenv.2022.109469.

Zhang, Jun, Peng Cui, and Haihong 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 (August): 107035. https://doi.org/10.1016/j.buildenv.2020.107035.

Zhang, Yan, Bak Koon Teoh, Maozhi Wu, Jiayu Chen, and Limao Zhang. 2023. "Data-Driven Estimation of Building Energy Consumption and GHG Emissions Using Explainable Artificial Intelligence." Energy 262 (January): 125468. https://doi.org/10.1016/j.energy.2022.125468.

Zheng, Xing, Liutao Chen, and Jiachuan 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 (January): 109815. https://doi.org/10.1016/j.buildenv.2022.109815.