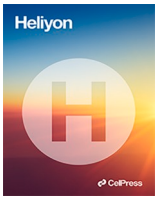
Heliyon 10 (2024) e29137

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Data-mining framework for in-depth quantitative study of influences on low-wind-velocity area from morphological parameters of cuboid-form buildings

Han Guo [a,b,d], Yehao Song [c], Yingnan Chu [c], Yi He [d,e,*], Weizhi Gao [c], Xiaoqing Guan [c]

[a] Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources of China, Shenzhen, China
[b] School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
[c] School of Architecture, Tsinghua University, Beijing, China
[d] School of Architecture and Urban Planning, Huazhong University of Science and Technology, Wuhan, China
[e] National Center of Technology Innovation for Digital Construction, Wuhan, China

ARTICLE INFO

Keywords:
Wind environment
Data mining
Quantitative analyses
Hybrid model

ABSTRACT

Wind environment is important in architectural sustainable design, as existing studies show that it can be considerably influenced by building morphologies. This study aimed to develop a data-mining framework to quantitatively evaluate and compare influences on Low-Wind-Velocity Area (LWVA) of common cuboid-form buildings with typical morphological parameters. The data-mining framework was originally developed by integrating multiple computational methods for rapid in-depth iterative analyses, including the generation of building models using parametric modelling, the big data generation based on hybrid model, and the statistical metric analysis method. The hybrid model was created by combining the CFD model and machine learning model. The accuracy and efficiency of the framework were fully demonstrated through the comprehensive validation and analyses of different models. The data of more than fifty thousand building cases with different morphological parameters and relevant wind conditions were generated and analyzed. Influences on LWVA of morphological parameters of cuboid-form building was comprehensively evaluated, including the visualization of multiple parameters, calculation and comparison of several correlation coefficients. It suggested that the reduction of height and width on the windward side would significantly decrease the LWVA and promote the outdoor ventilation. The change of depth would have relatively limited influence on the LWVA. Multivariate regression model-fit and variance analyses were further implemented, and it was found that there was a relatively significant linear correlation between the LWVA and morphological parameters. The equation of multivariate regression model was provided for extra rapid prediction. The study outcome could contribute to efficient evaluation of LWVA and provide useful information for sustainable design.

* Corresponding author. School of Architecture and Urban Planning, Huazhong University of Science and Technology, Wuhan, China.
E-mail addresses: guohan@whu.edu.cn (H. Guo), ieohsong@mail.tsinghua.edu.cn (Y. Song), zhuyn18@mails.tsinghua.edu.cn (Y. Chu), yihe@hust.edu.cn (Y. He), gwz20@mails.tsinghua.edu.cn (W. Gao), gxq22@mails.tsinghua.edu.cn (X. Guan).

https://doi.org/10.1016/j.heliyon.2024.e29137
Received 1 February 2024; Received in revised form 31 March 2024; Accepted 1 April 2024
Available online 4 April 2024

1. Introduction

Under the background of global warming, wind environment is gaining importance in architectural sustainability [1]. It is closely related to urban heat island, air renewal, thermal comfort, and building energy usage. Nowadays, wind environments influenced by building morphologies are often considered in the relevant studies of urban micro climates and architectural design in the early stage [2]. Comparison of numerous designs would help in the exploration of underlying mechanisms and design optimizations, especially for multiple morphological parameters. It is relatively time-consuming and complicated to implement iterative analyses by using conventional methods with repetitive modelling and complex simulations [3]. Comprehensive evaluations for multiple building morphologies are relatively primitive. These motivate us to develop a novel methodology based on data mining approach for in-depth quantitative analyses of wind environments.

1.1. Conventional methods and buildings in wind-environment studies

Experimental and Computational Fluid Dynamics (CFD) methods were widely used to evaluate wind environments, wind energy usage, thermal comfort and built environmental under certain conditions [4]. They were demonstrated by numerous studies and engineering applications [5,6]. Experimental and CFD methods often required input wind data obtained from measurement [7]. There were usually strict requirements for the measurement in many aspects. These often made the entire process slow, complicated, and expensive. Therefore, prediction technologies with low cost were becoming an optional efficient choice [8]. Hao et al. thought that existing studies of outdoor thermal comfort and behavior are limited by the so-called 'small data' approach [9]. Most studies employed site measurement, questionnaires and interviews in the last two decades. Those studies were labor-intensive and the methods mentioned above could only cover limited sites over a particular period of time. Because of the limitation of sample size, the 'small data' approach was underpowered statistically. As the data collection protocols were often different in different studies, it was difficult to compare the findings from different studies.

Cuboid form was one of the most common forms for buildings. We implemented an on-site investigation of 383 residential communities in an urban area of China. And cuboid-form buildings could be found in 221 residential communities. Many kinds of buildings were in cuboid form with different sizes, including residential buildings, office buildings, commercial buildings, educational buildings, and so on. Aynsley et al. used to call the kind of buildings as 'bluff bodied' [10]. Winds would be separated at the windward surfaces after meeting the cuboid-form buildings [11]. Wind tunnel experiments suggested that the patterns of wind velocity magnitudes around buildings were complicated and influenced by building form ratios and wind directions [12].

1.2. Relevant methods of big data and artificial intelligence

Nowadays, with the development of big data and artificial intelligence, usages of the cutting-edge technologies in relevant studies of built environment kept on increasing. Zhang et al. used machine learning model to predict environmental indicators via morphological indicators. They compared seven machine learning algorithms for modelling the nonlinear relationship between the building morphology and the outdoor environments of 150 workers' villages in Shanghai [13]. Zhao et al. employed a deep learning simulation method to explore the effects of land use types and density on the spatial distribution of PM2.5 pollutants in the city of Wuhan [14]. Kabošová et al. introduced and tested an environment-driven design technique at the urban and architectural scale. The optimal design solution was merged for the urban configuration and architectural shape by utilizing the interplay between the architectural intention and weather influences. They did a case study through the real-time iterative analysis of the environmental performance of multiple design variants [15]. Li et al. developed a deep transfer learning neural network to predict and simulate more accurate and realistic building energy usage. An on-campus experiment with a wireless sensing system was conducted. The accuracy of local microclimate condition prediction could be improved by using the proposed method [16]. Hao et al. used social media to quantify park attendance in response to hot weather conditions. They collected a lot of geographically coded Twitter data in a large urban park in the city of Hong Kong in different climates. They acquired park attendance data, and captured occupant thermal sensation and comfort using a questionnaire. The performance of biometeorological indicators was compared [9]. There were some relevant studies of wind environments and urban micro climates. Lee et al. proposed a machine learning algorithm by considering the terrain features from satellite imageries to overcome the limitation of engineering judgment [17]. They compared effects of terrain similarity and distance from station on basic wind speed. They claimed that it should be the first time to explore this kind of artificial intelligence approaches to determine the basic wind speed.

There was a great potential for the usage of artificial intelligence systems to assess predictive modifications of urban wind, including velocities, energy, loading and so on. Khattak et al. used wind tunnel experiments combined with interpretable tree-based machine learning algorithms to estimate turbulence intensity near airport runways [18]. Glumac et al. investigated a multi-fidelity machine learning framework by taking advantage of the main benefits of two CFD approaches to ensure the simulation accuracy of wind loading while maintaining the computational efficiency [19]. Wu et al. proposed a fast courtyard wind simulation platform based on the parallel courtyard Lattice Boltzmann Method to optimize the design structure and wind environments [20]. Higgins et al. used artificial intelligence tools to generate an adequate database to improve the generation of urban wind energy. Their study presented the wind tunnel experimental results of buildings with different forms [21].

Some studies focused on the development and evaluation of relevant methods using the cutting-edge technologies. Park et al. compared performances of eight machine learning algorithms for predicting natural ventilation rate. Their study found that the algorithm of deep neural network has the best prediction performance with the evaluation metrics [22]. Zheng et al. proposed a

framework and a deep learning model for urban local dense wind speed forecasting. They developed a Convolutional Long Short-Term Memory (ConvLSTM) and Long Short-Term Memory (LSTM) combinatorial deep learning model to learn the features of input historical weather image series of Hong Kong datasets. Their proposed model was verified, and could effectively forecast wind speed series with large amplitude and rapid frequency changes [23].

1.3. Research gaps

Though conventional simulation methods such as CFD models have been widely recognized, they are often complicated in simulation settings and their computational running time is relatively long. Especially, comparative investigations that require iterative analyses would be time-consuming. Emerging computational methods have powerful functions in morphological modelling, predictions and evaluations. It would be important to explore their potential in sustainable design and wind-environment studies.

Research gaps could be summarized as follows: (1) In-depth quantitative analyses of building wind environments influenced by morphological parameters and consideration of relevant sustainable design strategies were not many. (2) Application of data mining and integration of multiple computational models was primitive in studies of building wind environments. (3) Existing studies could provide limited information to wind-environment optimization based on comprehensive adjustment of morphological parameters in large ranges.

In our study, these research gaps will be covered from several aspects: (1) In-depth quantitative analyses using computational methods: A data-mining framework will be proposed by integrating multiple computational methods, including parametric modelling, hybrid models, and statistical analyses, to provide efficient, accurate, and in-depth analyses of Low-Wind-Velocity Area (LWVA) influenced by building morphological parameters. This covers the gap by presenting a novel approach that utilizes computational methods for comprehensive evaluation. (2) Application of data mining and integration of multiple computational models: after the development of the data-mining framework and demonstration of its accuracy and efficiency, the study covers the gap by introducing an advanced methodology for wind environment studies. (3) Limited information for wind-environment optimization: The study provides specific findings regarding the influences of morphological parameters on LWVA, quantitatively evaluating their effects. It also offers a multivariate regression model for rapid prediction of LWVA. By providing detailed insights into how morphological parameters affect LWVA and offering a predictive model, the study addresses the gap by providing valuable information for wind-environment optimization in sustainable design.

1.4. Research framework

Based on the analysis of research gaps, this study will focus on the development of data-mining framework that can quantitatively evaluate and compare influences on the Low-Wind-Velocity Areas (LWVAs) of typical morphological parameters of common cuboid-form buildings. The LWVAs have been defined as the areas with the wind velocities under the product of 0.3 and the original reference wind velocities. The value of 0.3 was the wind velocity ratio provided by assessment criteria of wind environment at the pedestrian level (height = 1.5 m) around buildings from previous studies [24,25]. As presented in Fig. 1 below, the data-mining framework will be developed by integrating multiple computational methods to implement efficient in-depth iterative analyses for LWVAs influenced by the morphological parameters. As explained in the following methodology section, there will be three major parts in the development of the data-mining framework. First, 3D building models will be created automatically using parametric modelling. Second, big data of LWVAs related to numerous building cases will be generated based on the development of hybrid model. Third, the generated data will be evaluated by using the statistical metric analysis method for implementing the data mining.

In the creation of hybrid model, the CFD simulation model and machine learning model will work in coordination to efficiently generated big data for mining. Both the CFD model and machine learning model will be validated to ensure the accuracy of generated data. The LWVAs and morphological parameters will be visualized, and correlation coefficients will be calculated and compared comprehensively. Influences on LWVA of morphological parameters of cuboid-form buildings will be comprehensively evaluated,

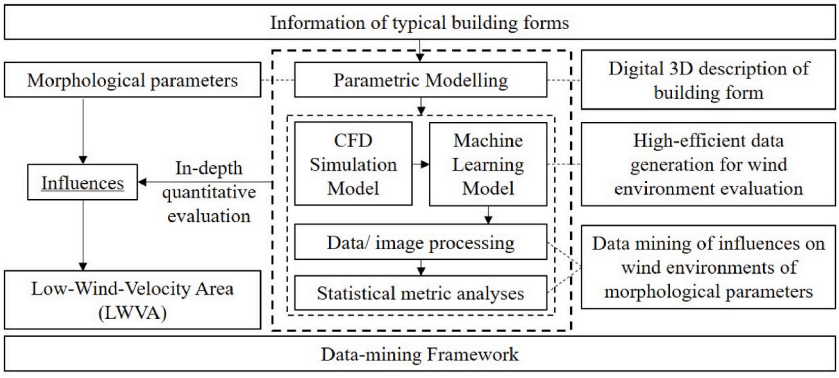


Fig. 1. The data-mining framework.

including the visualization of multiple parameters, calculation and comparison of several correlation coefficients. Multivariate regression model-fit analysis and variance analysis will be implemented to further analyze the linear relationship between the parameters. The multivariate regression model will also provide rapid prediction function after the validation. The study outcome will contribute to efficient evaluation and provide useful information for the optimization of LWVA in sustainable design.

2. Research methodology

2.1. Generation of 3D building models using parametric modelling

(1) Geometry and case settings of cuboid-form buildings

The establishment of hybrid model required massive data of LWVAs of building cases for training the machine learning model. The study intended to implement hundreds of cases of CFD simulations. LWVAs of cuboid-form buildings with various heights, widths and depths would be simulated under different wind velocities (Fig. 2). As presented in Table 1, there were 100 buildings cases with combinations of various morphological parameters; the inlet wind velocities on the reference height were set as 1 m/s, 2 m/s and 3 m/s for the 100 cases; there were totally 300 cases of CFD simulations as training data.

In the table, the values of widths, depths and heights were typical for common residential buildings in modern cities around the world; the values of wind velocities on the reference height could be universally observed in urban areas. In this study, wind velocities have been respectively set as 1, 2 and 3 m/s in the development of the hybrid model. These are typical wind velocities in urban areas, especially for the developing cities with relatively high densities in the middle south east of China. The wind conditions of simulations were set according to well-recognized climate data, including International Weather for Energy Calculation (IWEC), Chinese Typical Year Weather (CTYW) and Solar and Wind Energy Resource Assessment (SWERA). Take Wuhan City as an example, the local monthly average wind velocities are in the range of 1.0–1.5 m/s; the local monthly average high wind velocities are in the range of 2.0–3.0 m/s; the local monthly average low wind velocities are in the range of 0–1.0 m/s.

(2) Parametric modelling algorithm of cuboid-form buildings

Conventional manual modelling usually required drawings of lines and surfaces following several steps. The parametric modelling could be completed automatically using program based on the given morphological parameters of width, depth, and height. Similarly, the cuboid boundaries could also be generated using the other three parameters for the subsequent CFD simulations. It was efficient to generate numerous building models with different morphological parameters by simply adjusting parameters of the parametric-modelling program. In this study, the modelling scripts were developed in the environment of Grasshopper of Rhino [26]. In the program, the floor plane surfaces were created first based on the defined reference points; the cuboid-form building models were generated using the 'extruding' component with the surfaces and defined building heights.

2.2. Big data generation based on development of hybrid model

2.2.1. CFD simulation model

CFD simulation procedures consisted of governing equations, computational domain, and boundary conditions. All the simulations were following the well-recognized Best Practice Guideline COST 732 (European Cooperation in Science and Technology) [27].

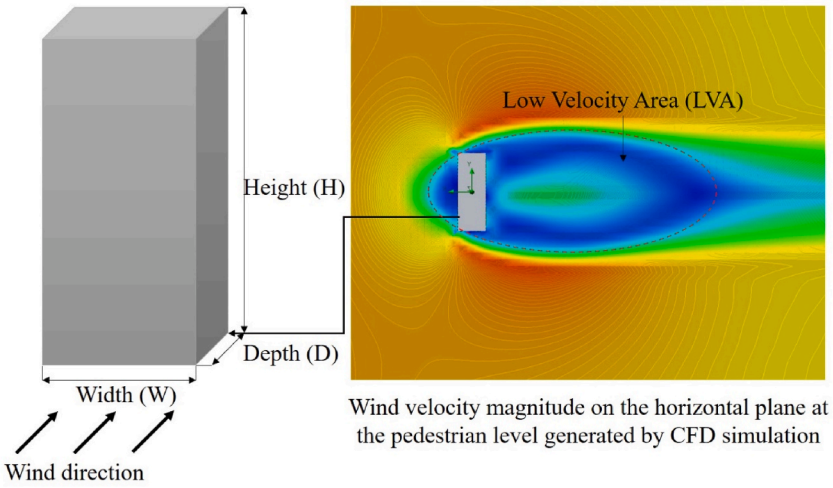


Fig. 2. The wind direction, morphological parameters and LWVA of cuboid-form building.

Table 1
Building cases with combinations of various morphological parameters.

Heights (m)	Widths (m)	Depths (m)
9, 12, 15, 18, 21	24, 30, 36, 42, 48	9, 12, 15, 18

Note: The inlet wind velocities on the reference height were set as 1 m/s, 2 m/s and 3 m/s for the 100 cases. There would be totally 300 cases of CFD simulations.

(1) Governing equations in CFD simulation

The RANS method demonstrated by existing studies and engineering applications was accurate and efficient for our study [26,24]. A modified k-ε model employed for the turbulent kinetic energy (k) and the dissipation rate (ε) was adopted for the simulation of 300 cases [28]. The CFD software Flow Simulation was adopted to solve the RANS equations. As presented in following transparent equations, ρ is the fluid density, μ is the dynamic viscosity; u_i and u_j are the velocities of the unit volumes x_i and x_j; σ_k and σ_ε are the turbulence Prandtl numbers corresponding to the turbulent kinetic energy and dissipation rate; C_μ, $C_{\varepsilon 1}$, and $C_{\varepsilon 2}$ are the empirical constants; $\mu_t = f_u \frac{C_\mu \rho k^2}{\varepsilon}$ is the turbulence viscosity; $P_B = -\frac{g_i}{\sigma_B}\frac{1}{\rho}\frac{\partial \rho}{\partial x_i}$, if $P_B < 0$, $C_B = 1$, if $P_B > 0$, $C_B = 0$; the model reverts to the standard k-ε model if the functions f_μ, $f_1 = 1 + \left(\frac{0.05}{f_u}\right)^3$ and $f_2 = 1 - e^{R_t^2}$ are equaling 1 [29].

$$\frac{\partial \rho k}{\partial t} + \frac{\partial \rho k u_i}{\partial x_i} = \frac{\partial}{\partial x_i}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_i}\right] + \tau_{ij}\frac{\partial u_i}{\partial x_j} - \rho\varepsilon + \mu_t P_B \tag{1}$$

$$\frac{\partial \rho \varepsilon}{\partial t} + \frac{\partial \rho \varepsilon u_i}{\partial x_i} = \frac{\partial}{\partial x_i}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_i}\right] + C_{\varepsilon 1}\frac{\varepsilon}{k}\left[f_1 \tau_{ij}\frac{\partial u_i}{\partial x_j} + C_B \mu_t P_B\right] - f_2 C_{\varepsilon 2}\frac{\rho \varepsilon^2}{k} \tag{2}$$

$$\tau_{ij}^R = \mu_t s_{ij} - \frac{2}{3}\rho k \delta_{ij} \tag{3}$$

$$s_{ij} = \frac{\partial u_i}{\partial x_j} + \frac{\partial u_i}{\partial x_i} - \frac{2}{3}\delta_{ij}\frac{\partial u_k}{\partial x_k} \tag{4}$$

(2) Computational domain, grid generation and boundary conditions

All the simulation cases were set with the same size of computational domain following the guidelines COST 732 [27] (Table 2). Grids of CFD simulations were generated according to parametric building models by adopting Cartesian-based technology. The generation procedures included the subtraction of building models using body-fitting algorithm, division of computational domain, and automatic refinement [30]. Grids were simplified with appropriate densities to promote simulation efficiency and reduce computational loads.

The reference wind velocities on the 10 m height were set as 1.0 m/s, 2.0 m/s and 3.0 m/s in CFD simulations. The initial wind velocity on the vertical profile can be defined by the equation below: u_{ABL}^* is the reference velocity, κ is the von Karman constant, z is the height, z_0 is the aerodynamic roughness longitude [31].

$$U(z) = \frac{u_{ABL}^*}{\kappa}\ln\left(\frac{z + z_0}{z_0}\right) \tag{5}$$

2.2.2. Data/image processing method

Data/image processing method was used for the quantitative comparison of the LWVAs at the pedestrian level (height = 1.5 m) around buildings [32]. In this study, the analysis boundaries were all in the same sizes as the simulation boundaries. Wind-velocity-magnitude images with the same sizes and resolutions were processed in the image processing tool of ImageJ. Slight differences among CFD simulation results could be distinguished based on accurate pixel value statistics. In the images, a pixel was the smallest area with a given color according to the numerical value of wind velocity at a coordinate. The single-pixel area could be

Table 2
Size of computational domain for all the simulation cases.

Windward extension	120 m
Leeward extension	330 m
Lateral extension	150 m
Vertical extension	150 m

Note: Extensions are between the building and domain boundaries.

calculated based on the setting scale of the boundary width (450m-1516 pixels). A LWVA could be obtained by multiplying the single-pixel area (0.088 m^2) and measured pixel number of the area with particular colors.

2.2.3. Machine learning model

(1) Creation of machine learning model

The script of machine learning model was created in the software program Jupyter Notebook of Anaconda. Several libraries including Numpy, Pandas, and Scikit-Learn were adopted to provide two major functions of prediction and prediction evaluation. The model of 'Extremely Randomized Trees' was selected due to its relatively high accuracy and efficiency for predictions of LWVAs in this study. Its functions included classification, regression and clustering [33]. The model could improve generalizability or robustness as it reduced the variance by combining predictions of several base estimators. Its randomness went a further step in computing the way of splits in comparison with the similar averaging algorithm of Forests of Randomized Trees. In addition, the best threshold was selected from all the thresholds generated for every candidate feature based on the splitting rule. Because the variance of prediction model could be further reduced in comparison with Forests of Randomized Trees [33].

(2) Training and improvement of machine learning model

The Excel files that contained building information and analysis results of LWVAs were read and standardized. The irrelevant, duplicate or incomplete data were removed or modified, and text-based data were converted to numerical values; the data were further processed by dropping the morphological parameters of cuboid-form buildings as the original parameters of features; the rest parameters of LWVAs were used as the prediction regression target for the original parameters.

The entire data set was split into two segments of training data and test data for the machine learning model. The proportion of training data and test data was decided by setting the parameters appropriately using the algorithm with the split function of Pandas library. The training data were fed to train the machine learning model using the 'Extremely Randomized Trees' algorithm. The model would look for the pattern in the data automatically to make the original parameters of features fit the prediction regression target. After the training, the initial machine learning model could make primary predictions of LWVAs based on the given building morphological parameters. The predictions were evaluated and the parameters of the machine learning model were adjusted for several times to improve the accuracy of results.

2.2.4. Validation method of hybrid model

(1) Grid-independence analysis for CFD simulation model

The Grid Convergence Indexes (GCIs) were calculated in the grid independence analysis for CFD simulation model. The calculation process is presented in the following equations: Fs is the safety factor ($Fs = 1.25$), ε_{rms} is the root mean square relative error, r is the grid refinement ratio ($r = 2$), p is based on the second-order discretization of all terms ($p = 2$), n is the number of measurement points ($n = 10$), ε_i is the value of velocity (V) or turbulence kinetic energy (TKE) of winds at measurement points [34,35]. Three sets of coarse grid, medium grid and fine grid were constructed for the calculation and analysis of GCIs.

$$GCI = F_s \frac{\varepsilon_{rms}}{r^p - 1} \tag{6}$$

$$\varepsilon_{rms} = \left(\frac{\sum_{i=1}^{i=n} \varepsilon_i^2}{n} \right)^{\frac{1}{2}} \tag{7}$$

(2) Experimental validation for CFD simulation model

CFD simulation model was further validated by comparing its result with a well-recognized wind-tunnel experiment done by Brown et al. from the US Environmental Protection Agency [26,36,37]. In the wind tunnel, total 77 cubes were placed in 11 rows and 7 columns. The wind velocities on the 12 profiles in the centerline of the cube array were compared through the evaluation of statistical discrepancies between the experiment and CFD simulation [38,39]. The calculation method of statistical discrepancies is presented in the following subsection.

(3) Calculation of R-squared for machine learning model

R-squared was calculated to evaluate the prediction performed by machine learning model. It was a statistical measure of fit which could indicate the variation between the predicted dependent variable and original parameters [40]. The calculation was implemented by using the metric method of SK-Learn. As shown in the equation below, u is the residual sum of squares, v is the total sum of squares [41].

$$R^2 = 1 - \frac{u}{v} \tag{8}$$

(4) Calculation of statistical discrepancies for hybrid model

Statistical discrepancies were analyzed for the validation of hybrid model, including the CFD simulation model and machine learning model. This study respectively evaluated the statistical discrepancies between the wind velocities generated by the experiment and CFD simulation, and the statistical discrepancies between the LWVAs predicted by CFD simulation and machine learning model. Predicted Root Mean Square Division (PRMSD), Normalized Mean Square Error (NMSE), Fractional Bias (FB), and Correlation Coefficient (CC) were calculated. Their calculation processes were presented in the following equations: n is the number of measurement points, X_i and R_i are two values measured at a point for comparison, X_m and R_m are the mean values of series of X_i and R_i respectively (X_i, X_m are the values for validation; R_i, R_m are the values for reference, E.g.: CFD simulation versus experiment, machine learning prediction versus CFD simulation).

$$PRMSD_{ES} = \frac{1}{E_m} \times \sqrt{\sum\nolimits_{i=1}^{n} \frac{(E_i - S_i)^2}{n}} \tag{9}$$

$$NMSE_{ES} = \frac{\sum_{i=1}^{n} (E_i - S_i)^2}{\sum_{i=1}^{n} (E_i \times S_i)} \tag{10}$$

$$FB_{ES} = \frac{E_m - S_m}{0.5 \times (E_m + S_m)} \tag{11}$$

$$CC_{ES} = \frac{\sum_{i=1}^{n} [(E_i - E_m) \times (S_i - S_m)]}{\sqrt{\sum_{i=1}^{n} (E_i - E_m)^2} \times \sqrt{\sum_{i=1}^{n} (S_i - S_m)^2}} \tag{12}$$

2.3. Statistical metric analysis method

Statistical metrics analysis method was adopted for the analysis of influences of building morphological parameters. Customized analysis programs were developed using R-project language to analyze the data features, including distribution, correlation and so on.

2.3.1. Visualization method of data distribution

The 'excel' files were created first to contain the data of LWVAs and morphological parameters of buildings. The unreasonable data were eliminated, and missing data were supplemented. Then, the data was imported and converted to matrices. The data were visualized in figures to show the distributions of parameters for comparisons. The plotting was implemented by adopting the 'gather' function and the component of 'ggplot' in the R-project environment. The arrangements of figures were improved by adjusting or adding the parameters of functions.

2.3.2. Statistical test method

The well-recognized Shapiro-Wilk test (S–W test) was selected as for normality test in the statistical test. The sampling distribution tended to be normal in large samples, regardless of the data shape [42]. As presented in the equation below, the calculation of S–W test is based on the correlation between the data and corresponding normal scores [43]: n is the number of observations, $x_{(i)}$ are the values of the ordered sample, a_i is the tabulated coefficients.

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{13}$$

2.3.3. Calculation method of correlation coefficients

The coefficients of Pearson, Spearman and Kendall Correlations were calculated for paired parameter groups respectively by using

Table 3
The ranges of correlation coefficients and correlation strengths.

Ranges of correlation coefficients	Strengths of correlation
0.0–0.2	Very weak
0.2–0.4	Weak
0.4–0.6	Middle
0.6–0.8	Strong
0.8–1.0	Very strong

the correlation calculation function with the three methods in the 'R Project' software tool. The relationship between the ranges of correlation coefficients and correlation strengths is presented in Table 3.

A Pearson Correlation Coefficient (PCC) was a statistical measure that could describe the strength of linear relationship between two groups of parameters [44]. The interval or ratio-level parameters should be bivariant distributed and linearly related. The PCC calculation process is presented in the equation below: N is the number of pairs of scores, x and y are parameters from two groups, $\sum xy$ is the sum of products of paired parameters, $\sum x$ and $\sum y$ are the sums of parameters of x and y respectively, $\sum x^2$ and $\sum y^2$ are the sums of squared parameters of x and y respectively.

$$\rho_p = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (\sum y)^2\right]}} \tag{14}$$

Because the parameters might not always meet the parameter requirements of PCC, Spearman Correlation Coefficient (SCC) and Kendall Correlation Coefficient (KCC) were calculated to further analyze the correlation comprehensively [45]. A SCC was another statistical measure that could describe the strength and direction of relationship between two groups of ranked parameters [46]. Analysis of KCC could provide a distribution free test of independence and a statistical measure of the strength of dependence between two groups of parameters [47]. The calculations of SCC and KCC are presented in the following two equations: (1) d_i is the difference in ranks of the parameters, n is the number of data points of the two parameters; (2) n_c is the number of concordant, n_d is the number of discordant.

$$\rho_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{15}$$

$$\tau = \frac{n_c - n_d}{n(n-1)/2} \tag{16}$$

2.3.4. Multivariate regression analysis and variance analysis

Multivariate Regression Analysis (MRA) aimed to analyze the linear correlation relationships between the independent and dependent variables of LWVAs and morphological parameters of cuboid-form buildings. Change pattern of LWVAs depended on the changes of morphological parameters could be described by the regression model. The strength of the linear correlation was estimated by fitting a line to the observed data. The analysis of significance, single stratum of parameters and covariance could be carried out [48, 49]. Variance Analysis (VA) aimed to obtain information about the quality, validity and performance of the multiple linear regression models. The factors which have most influences could be found in a multidimensional model. The variances within a model were split into several parts, and the relationship within them was set up [50].

3. Results analysis

3.1. Validation results of hybrid model

(1) Results analysis of GCIs

In the grid-independence analysis, a CFD simulation of a cuboid-form building was used as a case study. The building was 36 m in width, 12 m in depth, and 15 m in height. The inlet wind velocity was set at 2.0 m/s on the reference height. Coarse, medium and fine grids were constructed respectively. V and TKE of winds were measured at two lists of ten points with different heights on the centerline in front of the windward side (x = 30 m) and behind the leeward side (x = −50 m) of the building respectively.

As presented in Table 4, all the results of GCIs were less than 5% that could meet the recommendation from former studies [51]. The results proved the grid accuracy of the CFD simulation model. The model could predict the winds well using the three grid sets. The coarse grid set was selected to promote the simulation efficiency.

(2) Result analysis of R-squared

For the machine learning model, initial calculation results of R-squared were around 80%–95% in the stage of evaluation and improvement of the model. The R-squared could reach 97.7% when the adjustment parameters of test size equaled 0.1 and random state equaled 33 in the algorithm of 'Extremely Randomized Trees'. The result analysis showed that the R-squared was quite close to

Table 4
GCIs of V and TKE at the discrete measurement points.

Grids	Discrete points	Grid 1-2	Grid 2-3	Grid 1-3
GCI of V (%)	X = −50 m	0.995	4.457	1.378
	X = 30 m	2.607	1.440	3.875
GCI of TKE (%)	X = −50 m	2.509	1.874	4.227
	X = 30 m	3.311	0.946	4.244

100%. The relatively high R-squared demonstrated that the prediction was sounding correct.

(3) Results analysis of statistical discrepancies

Analysis of statistical discrepancies included the CFD simulation model and machine learning model of the hybrid model. In the evaluation of statistical discrepancies between the LWVAs predicted by CFD simulation and machine learning model, there were total 8 building cases with 3 different wind conditions (Table 5). For the 8 building cases: the morphological parameters were randomly selected within the entire range between the maximum and minimum morphological parameters of the training data; they were totally different from the training data of machine learning model.

The calculation results of statistical discrepancies are presented in Table 6, including the comparison of experiment and CFD simulation model, and the comparison of CFD simulation model and machine learning model. Results can be summarized as follows for the both two comparisons: (1) *PRMSD*, *NMSE,* and *FB* were relatively close to 0; (2) *CC* was close to 1. They could meet the standards recommended by the COST criteria: $NMSE < 1.5$, and $-0.3 < FB < 0.3$ [26,52]. The analysis of statistical discrepancies further suggested that: the prediction made by CFD simulation model had a good agreement with the experimental measurement; and the predictions made by the validated CFD simulation model and machine learning model were relatively close to each other. The accuracy of the hybrid model was demonstrated.

3.2. Visualization and correlation analysis of LWVAs and morphological parameters

The analysis focused on the correlation of LWVAs and morphological parameters. Total 52,500 building cases with different morphological parameters and relevant wind conditions were analyzed. The morphological parameters include W, D and H. A customized R-Project calculation program was applied to implement the statistical test, calculation of PCCs, SCCs and KCCs. The result of S–W test of W, D, H, and LWVA is presented in Table 7. The distribution of H, W, V, D, and LWVA of 52,500 building cases is presented in Fig. 3.

The statistical test results of all the parameters suggested that all the parameters were in non-normal distribution. The calculation results of PCCs, SCCs, and KCCs of LWVAs and morphological parameters (Table 8) showed that: the correlation of LWVA and building height was the highest, the second one was the correlation of LWVA and building width, the third one was the correlation of LWVA and building depth, and the lowest one was the correlation of LWVA and wind velocity (Fig. 4).

3.3. Multivariate regression analysis results and evaluation

(1) Multivariate regression model-fit analysis

In the multivariate regression analysis, the morphological parameters V, H, W, and D were used as the predictor variables; the LWVA was used as the response variable. As presented in Table 9, the coefficients, significance, multiple and adjusted R-square values, and p-value are generated in the multivariate regression model fit summary. The analysis shows that the morphological parameters V, H, W, and D are highly significant in model fitting. The p-value (<2e-16) was relatively low. Both the multiple and adjusted R-square reached 94.91% after adjustments; as they are close to 100%, they are relatively high. The analysis confirms the prediction accuracy of multivariate regression model. The visualization of coefficients of multivariate regression model are presented in Fig. 5. According to the coefficients, the LWVA prediction equation can be expressed as follows. A rapid prediction of LWVA can be easily implemented using the equation.

$$LWVA=(-37.71)\times V+(293.20)\times H+(107.85)\times L+(8.23)\times W-4123.65$$

In Fig. 6, residuals versus fitted, normal Q-Q, scale location, and residuals versus leverage are presented respectively. The results can be summarized as follows: (1) The residuals versus fitted plot can illustrate the non-linear relationship between the response and predictors. In this study, it shows that there are some nonlinear patterns in the residuals at the right according to the horizontal trend line. (2) The normal Q-Q plot shows that most the residuals can follow a straight dashed line. This suggests that the residuals are normally distributed. And there can be an assumption of linear regression. There are relatively small deviations when the theoretical

Table 5
The eight building cases with three different wind velocities.

Building case No.	Widths (m)	Heights (m)	Depths (m)
1	47	21	17
2	25	20	16
3	30	19	11
4	45	16	11
5	20	14	9
6	44	13	10
7	27	11	16
8	21	10	10

Note: Inlet wind velocities on the reference height were set as 1 m/s, 2 m/s and 3 m/s for the 8 building cases respectively.

Table 6
Results of PRMSD, NMSE, FB and CC for the validation of hybrid model.

	PRMSD	NMSE	FB	CC
Experiment & CFD simulation	0.159	0.026	0.104	0.972
CFD simulation & Machine learning prediction	0.087	0.312	0.067	0.977

Note: PRMSD-Predicted Root Mean Square Division; NMSE-Normalized Mean Square Error; FB-Fractional Bias; CC-Correlation Coefficient.

Table 7
S–W test results of LWVAs and morphological parameters.

Parameters	V	H	W	D	LWVA
W	0.63662	0.93536	0.95124	0.93536	0.985239
p-value	p-value <2.2e-16				

quantiles are less than −2.5 or passes 2.0 along the x-axis. (3) The scale location plot shows that there is a horizontal line with randomly spread points. This indicates good homoscedasticity. (4) The residuals versus leverage plot can show the relationship between the standardized residuals and leverage points of the observations in the regression model. Leverage points usually have a large influence on the model estimates. Residuals represent the differences between the observed and predicted dependent variable values. In this study, the residuals versus leverage plot presents that most standardized residuals are in a range of −2 to 2. There are a relatively few outlying values at the left and right in a range of 0.000025 and 0.00020. It suggests that a relatively few cases influence the regression line.

(2) Variance analysis

The result of variance analysis further demonstrated that the parameters of V, H, W and D had strong influences on the multivariate linear model (Table 10). The Df was the degrees of freedom for the independent variable and residuals. It referred to the number of independent values of parameters that could vary freely without breaking any constraint. The degrees of freedom were 1 for the four parameters of V, H, W and D. The Sum Sq was the sum of squares which were the total variation between the group means and the overall mean. The Mean Sq was the mean of the sum of squares, calculated by dividing the sum of squares by the degrees of freedom for each parameter. As the degrees of freedom were 1, the Sum Sq and Mean Sq were same for the four parameters respectively. It could be found that the F values were relatively large, especially for H and W (F value = 527582.54 for H; and F value = 449911.20 for W). The larger F value indicated that it is more likely for the variation caused by the independent variable is real and not due to chance. This is aligned with the correlation analysis: the morphological parameters of H and W had larger influenced on LWVA in comparison with V and D. The Pr (>F) is the p value of the F statistic. According to the p values (<2.2e-16), there were strong linear correlation between the LAV and the parameters of V, H, W, and D.

4. Discussion

4.1. Efficiency and accuracy of hybrid model

The parametric modelling could largely promote the modelling efficiency for numerous building cases. Parametric modelling could save more than 75% time in comparison with conventional manual modelling in this study. The advantage would be even more significant for more cases. The integration of CFD simulation model, machine learning model and statistical metric analysis could largely promote the evaluation efficiency of LWVAs of cuboid-form buildings. The multivariate regression analysis further provided the linear model as a rapid manual calculation method for the prediction of LWVAs.

The accuracy of hybrid model was fully demonstrated by the validation results analysis: the accuracy of CFD simulation model was proved by the results of analysis of GCIs and statistical discrepancies; the accuracy of machine learning model was proved by the results of analysis of R-square and statistical discrepancies.

4.2. Results discussion

(1) Trends of influences of the morphological parameters

The trends of influences of the morphological parameters presented by the analyses of the three kinds of correlation coefficients were similar and consistent. The calculation results of PCCs, SCCs and KCCs all suggested that: height of cuboid-form building had the largest influence on the LWVA; the influence of width of the building was the second largest. There was a close relationship between the windward area of cuboid-form building and its influence on the LWVA. The influence of depth of the building was much smaller in comparison with the height and width. Therefore, the reduction of building height and width would significantly decrease the LWVA. It would be reasonable to control the building height for the reduction of LWVA, especially for high-density cities.

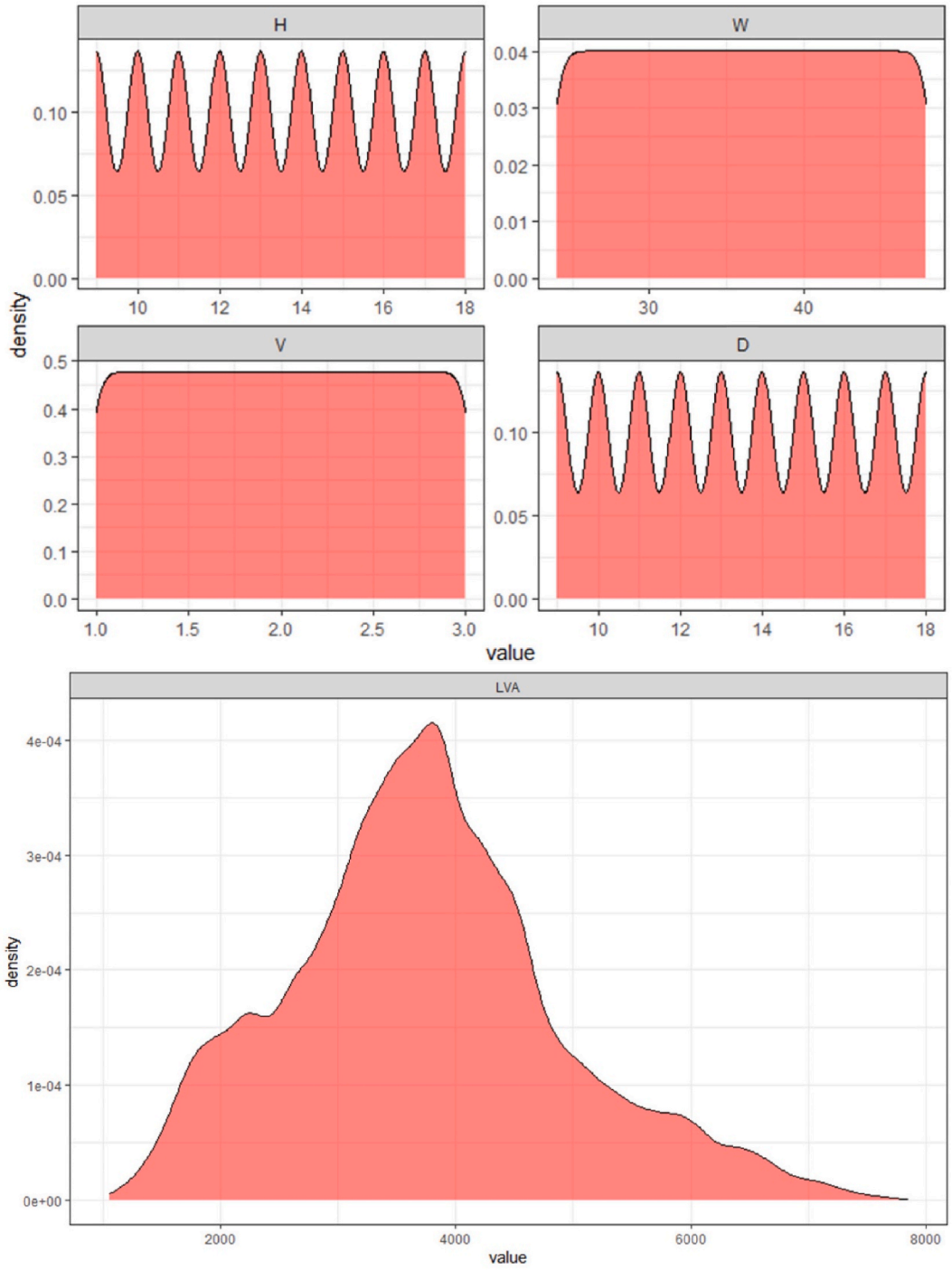


Fig. 3. Distribution of H, W, V, D, and LWVA of 52,500 building cases.

(2) Discussion of multivariate regression analysis

The multivariate regression model-fit analysis further proved that the morphological parameters of cuboid-form building could strongly influence the surrounding LWVA. The analysis clearly showed that there was a relatively significant linear correlation between the LWVA and morphological parameters. The multivariate regression model could be part of the hybrid model, as it was relatively easy to predict the LWVA by using the equation with the entering parameters.

Table 8
PCCs, SCCs, and KCCs of LWVAs and morphological parameters.

Parameters	Pearson correlation	Spearman correlation	Kendall correlation
Wind velocity	−0.019400	−0.017100	−0.011720
Height	0.715418	0.717054	0.541076
Width	0.660660	0.655217	0.479722
Depth	0.020086	0.019809	0.013672

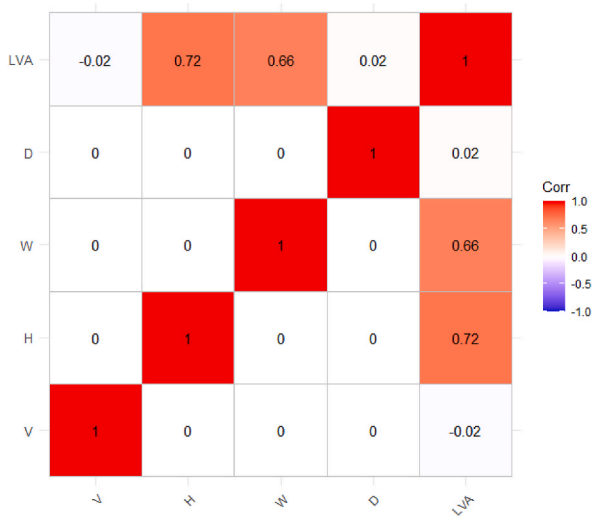


Fig. 4. Correlation-strength heat map of LWVAs and morphological parameters.

Table 9
A brief multivariate regression fit summary.

	Intercepts	V	H	W	D
Coefficients	−4123.65	−37.71	293.20	107.85	8.23
Significance	***	***	***	***	***
Multiple R-square		0.9491			
Adjust R-square		0.9491			
p-value		<2e-16			

Significance codes: 0–0.001: '***'; 0.001–0.01: '**'; 0.01–0.05: '*'; 0.05–0.1: '.'; 0.1–1: ''.

4.3. Research limitations

Though the hybrid model could promote the efficiency of LWVA evaluation, CFD simulation model could not be replaced by machine learning model. Fundamentally, CFD simulation models were based on physic models, and machine learning models were based on mathematical empirical models. Their working mechanisms were totally different. The development of machine learning model and its prediction function depended on the training data generated by the CFD simulation model. This study proved that the development of hybrid model could integrate the advantages from different computational methods.

In our study, hundreds of building cases with the morphological parameters in certain ranges were initially simulated and used as training data for the machine learning model. The number of building cases was still relatively few, especially for the development of some neural network models. This study only considered the particular cuboid form of buildings. However, realistic buildings would be in complicated and numerous forms and configurations. This study focused on the influences on low-wind-velocity area from the morphological parameters. The other factors related to wind environments have not been studied much. And this study only considered the wind direction that is vertical to the windward surface of the buildings. In the realistic situations, there could be numerous wind directions.

4.4. Potentials and future studies

There is a great potential that the methodology and similar workflow can be applied in relevant studies, building performance evaluation and sustainable design practices considering multiple complex factors. It would help in providing useful information for

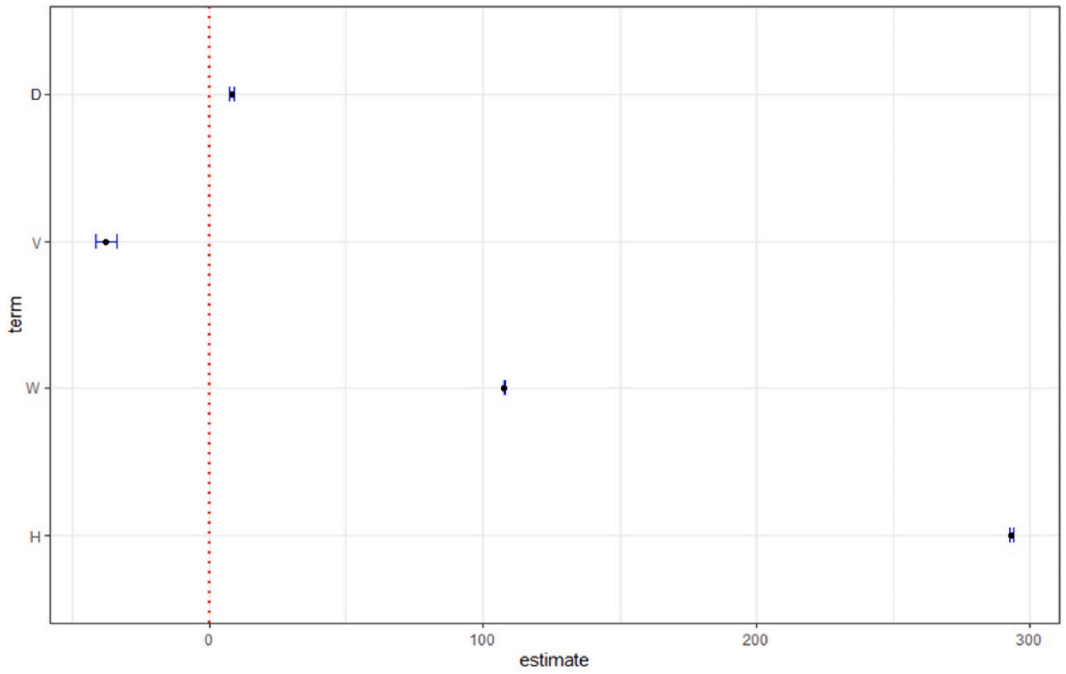


Fig. 5. Visualization of the coefficients of multivariate regression model.

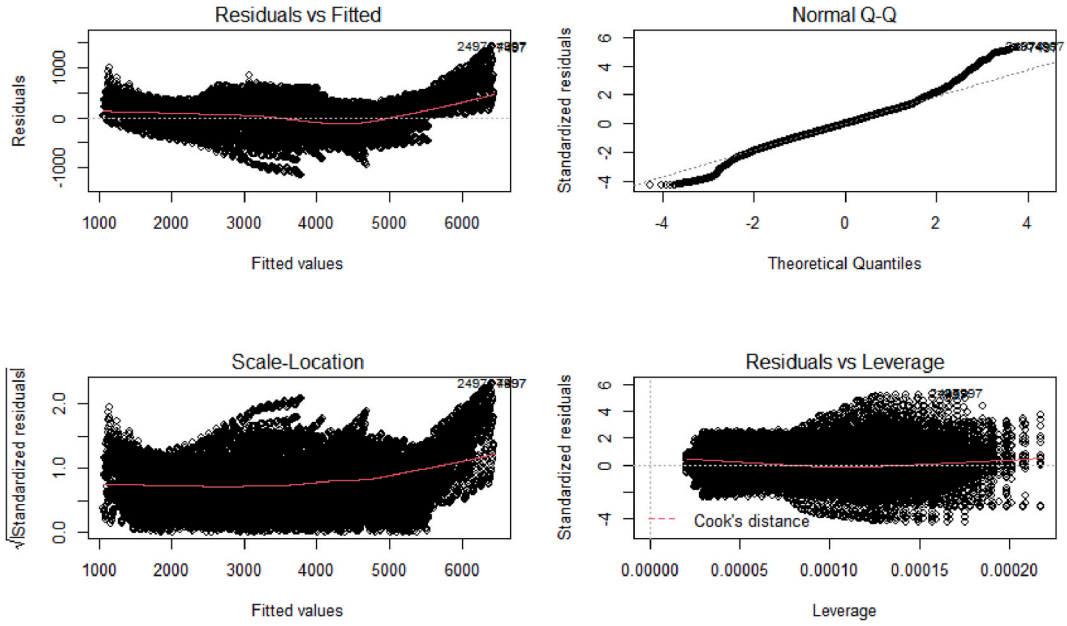


Fig. 6. Visualization of multivariate regression model of LWVAs prediction.

design strategy making and morphological optimization for buildings.

Buildings with the other representative forms and configurations can be studied in future research. Especially for influences on LWVAs of typical morphological parameters, similar customized models can be established for more in-depth analyses. And different wind directions and building orientations can be considered in the future studies. More relevant parameters of wind environments can be considered to comprehensively present the performance of building ventilation and urban micro climate. For example, there can be air exchange rate, wind velocities at particular points, and so on.

For the methodology, the models would be improved in different ways to further promote the efficiency and accuracy. Relevant algorithms can be developed to help in parametric modelling and simulation meshing. The other CFD simulation models such as LES

Table 10
A brief variance analysis summary.

Df	V	H	W	D
	1	1	1	1
Sum Sq	2.7371e+07	3.7234e+10	3.1753e+10	2.9351e+07
Mean Sq	2.7371e+07	3.7234e+10	3.1753e+10	2.9351e+07
F value	387.83	527582.54	449911.20	415.88
Pr (>F)	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16
Significance	***	***	***	***
Note	Df: Degrees of freedom; Sum Sq: Sum of squares; Mean Sq: Mean of the sum of squares; F value: Test statistic from the F test; Pr (>F): The p value of the F statistic.			

method can be tried with efficient computational resources. The machine learning models with a more precise prediction function can be tested. The integration of multiple models would be upgraded by inserting the physical equation in the modification of machine learning model.

5. Conclusion

This study has proposed a data-mining framework that can quantitatively evaluate and compare influences on Low-Wind-Velocity Area (LWVA) of common cuboid-form buildings with typical morphological parameters. The data-mining framework has been developed by integrating multiple computational methods for rapid in-depth iterative analyses, including the generation of building models using parametric modelling, big data generation using hybrid model, and statistical metric analysis method. The hybrid model has been created by combining the CFD model and machine learning model. Its accuracy and efficiency have been demonstrated through the comprehensive validation and analyses of different models. Influences on LWVA of morphological parameters of cuboid-form building have been comprehensively evaluated, including the visualization of multiple parameters, calculation and comparison of several correlation coefficients. Multivariate regression model-fit analysis and variance analysis have been implemented to further analyze and evaluate the linear relationship between the parameters. The equation of multivariate regression model has been provided for rapid prediction of LWVA, and its accuracy has been demonstrated. Specific findings can be summarized as follows.

(1) This study has presented the development process of a novel data-mining framework that can provide efficient, accurate and in-depth analyses of LWVA influenced by building morphological parameters. It is an original attempt to integrate the multiple computational methods of parametric modelling, hybrid model, and statistical metric analyses. The accuracy of the hybrid model based on CFD model and machine learning model has been validated by calculating the GCIs, R-squared and statistical discrepancies. The hybrid model can generate the big data of LWVA influenced by building morphological parameters rapidly and precisely. Small differences of numerous cases can be easily distinguished and compared. There is a great potential to adopt the framework in the other relevant studies and sustainable design optimization.
(2) Influences on LWVA of morphological parameters of cuboid-form building has been quantitatively evaluated: Influence of the height is the largest (PCC = 0.715, SCC = 0.717, KCC = 0.541); influence of the width is the second largest (PCC = 0.661, SCC = 0.655, KCC = 0.480); influence of the depth is much smaller than the height and width (PCC = 0.020, SCC = 0.020, KCC = 0.014). The reduction of height and width on the windward side would significantly decrease the LWVA and promote the outdoor ventilation. The change of depth would have relatively limited influence on the LWVA.
(3) There is a relatively significant linear correlation between the LWVA and morphological parameters of cuboid-form buildings. The LWVA can be calculated manually with the multivariate regression model by entering the initial wind velocity and morphological parameters of cuboid-form buildings in the provided equation.

Data availability statement

Data associated with this study has not been deposited into a publicly available repository. The data has been included in the article and references in this paper.

CRediT authorship contribution statement

Han Guo: Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **Yehao Song:** Supervision, Funding acquisition. **Yingnan Chu:** Resources, Data curation. **Yi He:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Weizhi Gao:** Software. **Xiaoqing Guan:** Visualization.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that inappropriately influence

our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Acknowledgements

This research was financially supported by the Natural Science Foundation of China (NSFC Grant No. 42371471; No. 52078264) and the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources of China (Grant No. KF-2022-07-022).

References

[1] Y. Toparlar, B. Blocken, B. Maiheu, G.J.F. Van Heijst, A review on the CFD analysis of urban microclimate, Renew. Sustain. Energy Rev. 80 (December) (2017) 1613–1640.
[2] B. Blocken, W.D. Janssen, T.V. Hooff, CFD simulation for pedestrian wind comfort and wind safety in urban areas: general decision framework and case study for the Eindhoven University campus, Environ. Model. Software 30 (2012) 15–34.
[3] He Yi, Xiao-Hui Liu, Hong-Liang Zhang, Wei Zheng, Fu-Yun Zhao, Aurel Schnabel Marc, Mei Yi, Hybrid framework for rapid evaluation of wind environment around buildings through parametric design, CFD simulation, image processing and machine learning, Sustain. Cities Soc. 73 (2021) 103092, https://doi.org/10.1016/j.scs.2021.103092. ISSN 2210-6707.
[4] Lang Zheng, Weisheng Lu, Qianyun Zhou, Weather image-based short-term dense wind speed forecast with a ConvLSTM-LSTM deep learning model, Build. Environ. 239 (2023) (2023) 110446, https://doi.org/10.1016/j.buildenv.2023.110446. ISSN 0360-1323.
[5] B. Blocken, T. Stathopoulos, J.P.A.J. van Beeck, Pedestrian-level wind conditions around buildings: review of wind-tunnel and CFD techniques and their accuracy for wind comfort assessment, Build. Environ. 100 (2016) (2016) 50–81, https://doi.org/10.1016/j.buildenv.2016.02.004.
[6] D.Y.C. Leung, Y. Yang, Wind energy development and its environmental impact: a review, Renew. Sustain. Energy Rev. 16 (1) (2012) 1031–1039, https://doi.org/10.1016/j.rser.2011.09.024.
[7] Hong Kong Observatory, Meteorological instruments –wind. https://www.hko.gov.hk/en/education/met_instrument/wind.htm?flash=1?flash=1. (Accessed 18 November 2022).
[8] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, Z. Yan, A review on the forecasting of wind speed and generated power, Renew. Sustain. Energy Rev. 13 (4) (2009) 915–920, https://doi.org/10.1016/j.rser.2008.02.002.
[9] Hao Tongping, Haoliang Chang, Sisi Liang, Phil Jones, P.W. Chan, Lishuai Li, Jianxiang Huang, Heat and park attendance: evidence from “small data” and “big data” in Hong Kong, Build. Environ. 234 (2023) 110123, https://doi.org/10.1016/j.buildenv.2023.110123. ISSN 0360-1323.
[10] R.M. Aynsley, Shape and flow: the essence of architectural aerodynamics, Architect. Sci. Rev. 42 (2) (1999) 69–74.
[11] Z.T. Ai, C.M. Mak, J.L. Niu, Numerical investigation of wind-induced airflow and interunit dispersion characteristics in multistory residential buildings, Indoor Air 23 (5) (2014) 417–429.
[12] W. Luo, Z. Dong, G. Qian, J. Lu, Wind tunnel simulation of the three-dimensional airflow patterns behind cuboid obstacles at different angles of wind incidence, and their significance for the formation of sand shadows, Geomorphology 139–140 (2012) 258–270, https://doi.org/10.1016/j.geomorph.2011.10.027.
[13] Xingzhao Zhang, Luqiao Yang, Ruizhe Luo, Hsin-Yu Wu, Jiaqi Xu, Chenyu Huang, Yingjun Ruan, Xiaowei Zheng, Jiawei Yao, Estimating the outdoor environment of workers’ villages in East China using machine learning, Build. Environ. 226 (2022) (2022) 109738, https://doi.org/10.1016/j.buildenv.2022.109738. ISSN 0360-1323.
[14] Liyuan Zhao, Ming Zhang, Si Cheng, Yunhao Fang, Shuxian Wang, Cong Zhou, Investigate the effects of urban land use on PM2.5 concentration: an application of deep learning simulation, Build. Environ. 242 (2023) (2023) 110521, https://doi.org/10.1016/j.buildenv.2023.110521. ISSN 0360-1323.
[15] Lenka Kabošová, Angelos Chronis, Theodoros Galanos, Stanislav Kmeť, Dušan Katunský, Shape optimization during design for improving outdoor wind comfort and solar radiation in cities, Build. Environ. 226 (2022) (2022) 109668, https://doi.org/10.1016/j.buildenv.2022.109668. ISSN 0360-1323.
[16] Li Qi, Wei Wang, Zhun Yu, Jiayu Chen, Assessing urban micro-climates with vertical and horizontal building morphological cutting deep transfer learning neural networks, Build. Environ. 234 (2023) (2023) 110186, https://doi.org/10.1016/j.buildenv.2023.110186. ISSN 0360-1323.
[17] Donghyeok Lee, Seung Yong Jeong, Thomas H.-K. Kang, Consideration of terrain features from satellite imagery in machine learning of basic wind speed, Build. Environ. 213 (2022) (2022) 108866, https://doi.org/10.1016/j.buildenv.2022.108866. ISSN 0360-1323.
[18] Afaq Khattak, Pak-wai Chan, Feng Chen, Haorong Peng, Estimating turbulence intensity along the glide path using wind tunnel experiments combined with interpretable tree-based machine learning algorithms, Build. Environ. 239 (2023) (2023) 110385, https://doi.org/10.1016/j.buildenv.2023.110385. ISSN 0360-1323.
[19] Anina Šarkić Glumac, Onkar Jadhav, Vladimir Despotović, Bert Blocken, Stephane P.A. Bordas, A multi-fidelity wind surface pressure assessment via machine learning: a high-rise building case, Build. Environ. 234 (2023) (2023) 110135, https://doi.org/10.1016/j.buildenv.2023.110135. ISSN 0360-1323.
[20] Renzhi Wu, Xiaoshan Fang, Shuang Liu, Qiong Li, Robert Brown, Junru Yan, A workflow for rapid assessment of complex courtyard wind environment based on parallel lattice Boltzmann method, Build. Environ. 233 (2023) (2023) 110112, https://doi.org/10.1016/j.buildenv.2023.110112. ISSN 0360-1323.
[21] Stéphanie Higgins, Ted Stathopoulos, Application of artificial intelligence to urban wind energy, Build. Environ. 197 (2021) (2021) 107848, https://doi.org/10.1016/j.buildenv.2021.107848. ISSN 0360-1323.
[22] Hansaem Park, Dong Yoon Park, Comparative analysis on predictability of natural ventilation rate based on machine learning algorithms, Build. Environ. 195 (2021) (2021) 107744, https://doi.org/10.1016/j.buildenv.2021.107744. ISSN 0360-1323.
[23] Lang Zheng, Weisheng Lu, Qianyun Zhou, Weather image-based short-term dense wind speed forecast with a ConvLSTM-LSTM deep learning model, Build. Environ. 239 (2023) (2023) 110446, https://doi.org/10.1016/j.buildenv.2023.110446. ISSN 0360-1323.
[24] Y. Li, L. Chen, Study on the influence of voids on high-rise building on the wind environment, Build. Simulat. 13 (2) (2019).
[25] Q.M.Z. Iqbal, A.L.S. Chan, Pedestrian level wind environment assessment around group of high-rise cross-shaped buildings: effect of building shape, separation and orientation, Build. Environ. 101 (5) (2016) 45–63.
[26] Y. He, M.A. Schnabel, Y. Mei, A novel methodology for architectural wind environment study by integrating CFD simulation, multiple parametric tools and evaluation criteria, Build. Simulat. (8) (2019) 1–17, https://doi.org/10.1007/s12273-019-0591-8.
[27] J. Franke, A. Hellsten, H. Schlünzen, B. Carissimo, Best practice guideline for the CFD simulation of flows in the urban environment. COST Office, 2007. Hamburg.
[28] B.E. Launder, D.B. Spalding, The numerical computation of turbulent flows, Comput. Methods Appl. Mech. Eng. 3 (1974) 269–289.
[29] C.K.G. Lam, K.A. Bremhorst, Modified form of model for predicting wall turbulence, ASME Journal of Fluids Engineering 103 (3) (1981) 456–460.
[30] A. Sobachkin, G. Dumnov, Numerical Basis of CAD-Embedded CFD, White Paper, 2014.
[31] P.J. Richards, R.P. Hoxey, Appropriate boundary conditions for computational wind engineering models using the k-ε turbulence model, J. Wind Eng. Ind. Aerod. 46 (1993) 145–153.
[32] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Methods 9 (7) (2012) 671–675.
[33] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (1) (2006) 3–42, 2006.
[34] L.F. Richardson, The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam, Phil. Trans. Math. Phys. Eng. Sci. 210 (1911) 307–357.
[35] P.J. Roache, Perspective: a method for uniform reporting of grid refinement studies, ASME Journal of Fluids Engineering 116 (1994) 405–413.

[36] M.J. Brown, R.E. Lawson, D.S. Decroix, R.L. Lee, Comparison of centerline velocity measurement obtained around 2D and 3D building array in a wind tunnel. Proceeding of the 2001 International Symposium on Environmental Hydraulics, 2001.
[37] S.J. Mei, J.T. Hu, D. Liu, F.Y. Zhao, Y.G. Li, Y. Wang, H.Q. Wang, Wind driven natural ventilation in the idealized building block arrays with multiple urban morphologies and unique package building density, Energy Build. 155 (2017) (2017) 324–338. ISSN 0378-7788.
[38] D. Wilks, Statistical Methods in the Atmospheric Sciences, second ed., Academic Press, Cambridge, 2006.
[39] J.C. Chang, S.R. Hanna, Air quality model performance evaluation, Meteorol. Atmos. Phys. 87 (1–3) (2004) 167–196.
[40] S.A. Glantz, B.K. Slinker, Primer of Applied Regression and Analysis of Variance, McGraw-Hill, 1990. ISBN 978-0-07-023407-9.
[41] N.R. Draper, H. Smith, Applied Regression Analysis, Wiley-Interscience, 1998. ISBN 978-0-471-17082-2.
[42] A. Ghasemi, S. Zahediasl, Normality tests for statistical analysis: a guide for non-statisticians, 2012 Spring, Int. J. Endocrinol. Metabol. 10 (2) (2012) 486–489, https://doi.org/10.5812/ijem.3505. Epub 2012 Apr 20. PMID: 23843808; PMCID: PMC3693611.
[43] Jiajuan Liang, Man-Lai Tang, Ping Shing Chan, A generalized Shapiro–Wilk W statistic for testing high-dimensional normality, 2009, Comput. Stat. Data Anal. 53 (11) (2009) 3883–3891, https://doi.org/10.1016/j.csda.2009.04.016. ISSN 0167-9473.
[44] E.R. Delong, D.M. Delong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (3) (1988) 837.
[45] Marie-Therese Puth, Markus Neuhäuser, Graeme D. Ruxton, Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits, Anim. Behav. 102 (2015) (2015) 77–84, https://doi.org/10.1016/j.anbehav.2015.01.010. ISSN 0003-3472.
[46] C. Spearman, The proof and measurement of association between two things, Int. J. Epidemiol. 39 (5) (2010) 1137–1150, https://doi.org/10.1093/ije/dyq191. October 2010.
[47] Inés Couso, Olivier Strauss, Hugo Saulnier, Kendall's rank correlation on quantized data: an interval-valued approach, Fuzzy Set Syst. 343 (2018) (2018) 50–64, https://doi.org/10.1016/j.fss.2017.09.003. ISSN 0165-0114.
[48] Sang Ku Park, Hyeun Jun Moon, Kyung Chon Min, Changha Hwang, Suduk Kim, Application of a multiple linear regression and an artificial neural network model for the heating performance analysis and hourly prediction of a large-scale ground source heat pump system, Energy Build. 165 (2018) (2018) 206–215, https://doi.org/10.1016/j.enbuild.2018.01.029. ISSN 0378-7788.
[49] Orkun Burak Öztürk, Ersan Başar, Multiple linear regression analysis and artificial neural networks based decision support system for energy efficiency in shipping, Ocean Eng. 2021 (2021) 110209, https://doi.org/10.1016/j.oceaneng.2021.110209. ISSN 0029-8018.
[50] Rakshit D. Muddu, D.M. Gowda, Anthony James Robinson, Aimee Byrne, Optimisation of retrofit wall insulation: an Irish case study, Energy Build. 235 (2021) (2021) 110720, https://doi.org/10.1016/j.enbuild.2021.110720. ISSN 0378-7788.
[51] S. Vinchurkar, P.W. Longest, Evaluation of hexahedral, prismatic and hybrid mesh styles for simulating respiratory aerosol dynamics, Comput. Fluid 37 (2008) 317–331.
[52] G. Efthimiou, J. Santiago, A. Martilli, COST 732 in Practice: the MUST Model Evaluation Exercise, 2011.