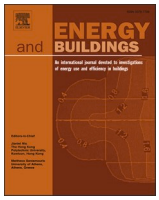
Energy & Buildings 344 (2025) 115974

Contents lists available at ScienceDirect

Energy & Buildings

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



A machine learning-guided semi-empirical model for predicting single-sided natural ventilation rates

Jung Min Han [a,*], Wentao Wu [b], Ali Malkawi [c]

[a] Department of Architecture and Architectural Engineering, Yonsei University, Seoul, Republic of Korea
[b] Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand
[c] Harvard Center for Green Buildings and Cities, Graduate School of Design, Harvard University, Cambridge 02138 MA, USA

ARTICLE INFO

Keywords:
Machine learning
Random forest
Pressure coefficient
Single-sided natural ventilation
Semi-empirical model

ABSTRACT

Most of the state-of-the-art natural ventilation models were developed for either single-sided, or cross ventilation mode, or buoyancy-driven ventilation. Natural ventilation (NV) of a single zone may vary between different modes in different seasons depending on the design and the operation of other building systems. This paper tailors the machine learning embedded semi-empirical models to predict the natural ventilation rate in a single zone. The process of model development consists of two parts: 1) semi-empirical model development for single-sided ventilation with a local context 2) machine learning driven component to accurately predict a specific lab condition. By taking a case study, the series of steps were taken to validate model accuracy with an estimated flowrate in given window operable areas. Firstly, the contextual inputs and localized wind speed as well as window models were investigated. Finally, we developed a machine learning model to predict the localized lab environment by using pressure sensor's data on façade. The random forest model was trained and fine-tuned to predict localized pressure coefficients (Cp). Over 75 % of the predicted values fall within the model's ± 1 standard deviation credible interval, demonstrating not only high predictive reliability but also suitability for integration into empirical ventilation models. These results highlight the model's potential as a robust input generator for semi-empirical frameworks with locally collected weather data, particularly in applications involving window operation control and site-specific model calibration.

1. Introduction

Natural ventilation is a passive technology to reduce building energy consumption [1]. Despite numerous studies on natural ventilation [2], quantifying natural ventilation rate remains challenging [3] as it involves the knowledge of complex mechanisms of fluid mechanics and heat transfer [4]. The traditional method estimating natural ventilation rate is based on the standard orifice flow equation that calculates airflow rate proportional to the square root of pressure difference between two openings on opposite walls to form cross ventilation [5]. In many buildings, internal walls, partitions and other configurations often hinders cross ventilation [6] and single-sided natural ventilation is prevalent [7]. However, for single-sided natural ventilation, the orifice flow equation might not be suitable to predict the airflow rate due to the different window types, bi-directional flow at a single opening [8], and complex airflow around the window [9].

Researchers have proposed several empirical models to calculate volume flowrates for single-sided natural ventilation [9]. Warren (1977) proposed a simple model to calculate the volume flow rate through openings on the same wall and the model considers only wind driven ventilation [10]. Larsen and Heiselberg (2008) incorporated buoyancy effect into a single-sided ventilation model [11]. Heiselberg and Sandberg (2016) found that the discharging coefficients of different window configurations [12] significantly affect the estimation of the volume flow rate. Wang and Chen (2012) developed single-sided natural ventilation models for different widow types [7]. Followed by that, Wang et al. (2015) expanded the models for six more window types and integrated stack effect into the models [13]. Pan et al. (2019) and Gough et al. (2020) developed advanced mathematical formula for the single-sided ventilation by considering urban impacts based on experimental data [14,15]. The semi-empirical model proposed in all the above research utilized pressure coefficient (C_p) to predict ventilation rate (Q). The C_p associated uncertainty is high [16]. The importance of C_p on performance-driven design was highlighted by conducting a sensitivity analysis in the use of building energy simulation [17,18]. However,

* Corresponding author.
E-mail address: jhan@yonsei.ac.kr (J.M. Han).

https://doi.org/10.1016/j.enbuild.2025.115974
Received 10 February 2025; Received in revised form 28 May 2025; Accepted 1 June 2025
Available online 2 June 2025

Nomenclature

Name	Description
C_p	Pressure coefficient, defined as the difference between the static pressure at a point on the building's façade and a reference point in the free stream, normalized by the dynamic pressure at the same reference point.
Q	Airflow rate through openings.
U_{ref}	Reference wind speed, typically measured far from the building to avoid the effects of the building and surrounding structures.
C_d	Discharge coefficient, a factor that adjusts the ideal flow equation to account for real-world conditions such as friction and turbulence.
A_{eff}	Effective opening area, the actual area of the opening adjusted for its shape and the nature of the flow through it.
ΔT	Temperature difference between the inside and outside of the building.H: Height of the opening.
Z_0	Neutral plane height from the ground, a reference level at which the pressure is equal inside and outside.
θ_w	Wind angle, the angle at which wind approaches the window or opening.

Kikuchi et al [19] highlighted the difficulties in determining C_ps.

C_p is defined as the difference of static pressure at a point on the building's façade and a reference point in the free stream, and then normalized by the dynamic pressure at the same reference point [20]. C_p can be determined with full-scale measurements [21], wind tunnel tests [22], and CFD approaches [23]. Hensen (1991) conducted experiments to develop empirical C_p models that vary according to wind conditions and are significantly affected by surrounding conditions [24]. The empirical models mainly consider wind pressure as the driving force to facilitate natural ventilation [25]. Full-scale experiments have the limitations to control the wind conditions in the free stream. Wind tunnel experiments are therefore used to determine C_p in controlled wind conditions. Studies show that wind tunnel experiments have limitations on predicting C_ps due to limited building shapes and surrounding configurations [26,27]. CFD simulations [28] were also used to study the influence of different wind incident angles on wind pressure coefficients. The research pointed out the importance to obtain pressure coefficient with both vertical and horizontal pressure coefficient gradients. However, it is computationally expensive to obtain C_p gradients for different buildings in real-world conditions.

To overcome limitations of experiments and CFD, machine learning (ML) [29] and deep learning (DL) models were developed to predict C_p on buildings' facades. Vrachimi, Melo, and Cóstola (2017) developed a ML model to predict C_p using wind-tunnel experiment data from Air Infiltration and Ventilation Centre (AIVC) database [30]. Bre, Gimenez, and Fachinotti (2018) trained neural networks using published C_p data from the Wind Engineering Information Center of the Tokyo Polytechnic University (TPU) [31]. Van Nguyen and De Troyer (2018) also proposed the C_p prediction methods using their own web-based datasets to train model [23]. To the authors' best knowledge, ML and DL models are not yet developed for real-world buildings under real-time climate conditions.

This study addresses these gaps by proposing a Machine Learning-Guided Semi-Empirical Model (ML-SEM) for predicting pressure coefficients and natural ventilation rates in a single-sided ventilation setup. The proposed method leverages localized sensor data, including wind direction, temperature, and humidity, collected from Harvard HouseZero's facade sensors and weather stations. By replacing the generic C_p component in a traditional SEM with a Random Forest-based ML model trained on real-time data, the framework improves predictive accuracy and adaptability. This integration not only minimizes reliance on assumptions and standardized C_p values but also reflects the true environmental behavior around specific buildings. Moreover, we assess the limitations of existing SEMs, particularly in their treatment of the neutral plane level, local wind speed attenuation, and contextual sheltering effects. These factors are systematically corrected using CFD-calibrated parameters, empirical modifiers, and ML-predicted C_p values, resulting in a more robust and context-sensitive model. This hybrid approach enhances the operational relevance of NV modeling for use in model predictive control, building automation, and environmental performance simulations. By addressing both theoretical limitations and practical application challenges, this research contributes a novel methodology for enhancing NV prediction accuracy using a data-driven approach. The ML-SEM framework is also scalable to a wide range of other buildings with available sensor data, offering a promising pathway for integrating empirical modeling with smart building technologies.

The goal of this study is to propose a real-time data-driven method to predict pressure coefficient and natural ventilation rate. The detailed objectives include: 1) review existing simplified semi-empirical NV models and evaluate the sensitivity of these models to pressure coefficients; 2) develop a machine learning method and algorithm to predict pressure coefficient and thus natural ventilation rate; 3) use real-time sensor data to predict pressure coefficients and ventilation rates of a real building and compare with other models. The novelty of this work to provide ML based airflow prediction framework using an existing semi-empirical model. It further shows the applicability of existing single-sided empirical models to real-world situations with C_p predictions using localized sensor data. This research demonstrates how localized sensor data can be effectively leveraged to predict context-specific pressure coefficients, thereby addressing a fundamental limitation in current ventilation models.

2. State-of-arts of semiempirical model development

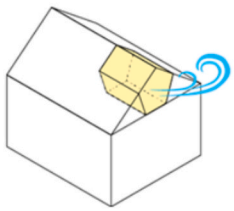
2.1. Overview of classical semi-empirical models for natural ventilation

Among existing models and methods, the level of simplicity of SEMs due to fewer parameters makes them useful for early design and system control. In this section, the simplified semi-empirical models (SEMs) are reviewed to estimate volume flow rate though openings of single-zone buildings. The results are further compared with CFD calculations in literature. Semi-empirical models were developed to calculate Q_n using a combination of parameters such as window's opening angle, wind speed, wind direction, height of neutral plane and the buildings (Fig. 1).

Several SEMs have been developed since the 1980s, including Warren (1977), ASHRAE (1987), Larsen & Heiselberg (2008), Wang (2015), and etc. Performance of some models are reviewed and compared in this section.

1) Warren (1977)

$$Q_n = f(\theta, \omega, d, C_p, P)$$



θ: window opening angle
ω: wind speed
d: wind direction
C_p: Pressure coefficient
P: Other fixed parameters

Fig. 1. Input parameters to calculate airflow rate.

$$Q_{stack} = \frac{1}{3} A_{eff} C_d \sqrt{\frac{\Delta T H g}{T_{ave}}} \tag{1}$$

$$Q_w = 0.025 A_{eff} U_w \tag{2}$$

where C_d is 0.6, A_{eff} is the effective opening area, Q_{stack} and Q_w are the flowrate for buoyancy and wind-driven ventilation.

2) ASHRAE Model (1987)

$$Q = C_v A U \tag{3}$$

where C_v = effectiveness of opening, and it is assumed to be 0.5–0.6 for perpendicular wind and 0.25–0.35 for diagonal winds.

3) Larsen & Heiselberg (2008)

$$Q = A \sqrt{C_1 f(\beta_v)^2 |C_p| U_{ref}^2 + C_2 H \Delta T + C_3 \left(\frac{\Delta C_p, opening \Delta T}{U_{ref}^2} \right)} \tag{4}$$

where ΔT is the temperature difference, H is the opening height, and the three constants C_1, C_2 and C_3 are found by use of the least squares method from their experiments.

This model took both wind pressure and temperature difference into consideration. They used a full-scale test room with controlled pressure and temperature to test the orifice equation. They found that the discharge coefficient cannot be regarded as a constant for airflow calculation due to different window types, opening area, and temperature difference across the opening.

4) Wang (2015)

$$Q = C(\theta_w, \alpha) \frac{C_d w \sqrt{C_p} \int_{z_0}^{h+h_0} \sqrt{z^{2/7} - z_0^{2/7}} dz}{z_{ref}^{1/7}} U_{ref} \tag{5}$$

where h is the height of the opening, h_0 is the elevation of the bottom of the opening above the ground and z_0 is the distance from the neutral level to the ground. where $C(\theta_w, \alpha)$ is the modifier. The equation for the modifier varies by the type of windows, as shown in Table 1.

Wang's model considers different types of windows such as hopper, awning, and casement windows. Wang's SEM was further validated by experiments in real environment. However, Wang's model did not consider temperature difference.

5) Pan (2019)

$$Q = C_d l \int_{z_0}^{h} \sqrt{\left| C_p \frac{U_{ref}^2}{z_{ref}^{2\alpha}} \frac{T_i}{T_\alpha} \left(z^{2\alpha} - z_0^{2\alpha} \right) + \frac{2(T_i - T_0)}{T_\alpha} g(z_0 - z) \right|} dz \tag{6}$$

where C_d is discharge coefficient, z is the height from ground, z_0 is the neutral plane height from ground, and T is air temperature. This empirical model considers wind speed, wind directions, and air temperature differences. The model has been applied to calculate buoyancy and wind-driven ventilation rate in apartment buildings. However, the above models do not consider the effect of flow fluctuation and the types of openings such as awning, hopper, and casement windows [32].

Table 1
Modifiers for different window configurations.

Window Type	$C(\theta_w, \alpha)$
Hopper	$min\left(\frac{C_{d,converging}}{C_{d,rec}} \sqrt{\frac{\sin\alpha}{2}}, 1 \right)$
Awning	$\frac{(1 - \lvert\cos\theta_w\rvert)}{2} \frac{w_1}{w_1 + \frac{w_2}{2}}$
Casement	$C(\theta_w, \alpha) = c \lvert\sin\theta_w\rvert \cos\frac{\alpha}{2}$ where $c = \begin{cases} 1 & 0 \le \theta_w \le 90 \\ 0.5 & otherwise \end{cases}$

Given that these constraints mentioned above, a structured model assessment is necessary to evaluate predictive accuracy under controlled geometry and environmental conditions. To this end, Section 2.2 presents a shoebox case study designed to quantify model error and propose targeted corrections to improve SEM reliability and contextual applicability.

2.2. Shoebox case study for model assessment and correction

While each of these semi-empirical models contributes valuable insights into airflow prediction under various conditions, their performance can vary significantly depending on window type, wind angle, and local environmental parameters. To critically assess their applicability and limitations in a controlled yet realistic setting, we implemented a comparative analysis using a shoebox building model. This setup allowed us to isolate key variables—such as window geometry, wind direction, and wind speed—and evaluate how well each model aligns with CFD benchmarks under consistent boundary conditions. The shoebox case serves as a practical testbed for quantifying the models' predictive accuracy and identifying the key parameters that require calibration for real-world applications. Building on this framework, we evaluated five SEMs alongside CFD results using the shoebox model proposed in Wang [7]. The shoebox building measured 3.6 m × 2.4 m × 3.3 m and included a 1 m × 1 m hopper window with a 30° opening angle. Among the five models tested, three—ASHRAE, Warren, and Wang—were selected for baseline comparison. These models were assessed against CFD results from Wu, Han, and Malkawi [33] to determine their predictive reliability.

Table 2 presents the calculated volume flow rates across a range of wind incident angles (0°–180°) and wind speeds (3–12 m/s). When wind was perpendicular to the window, the ASHRAE model predicted a high flowrate of 1.478 m^3/s, while Warren's model yielded 0.045 m^3/s, closely matching the CFD result of 0.049 m^3/s. Although Warren's model aligned well under this specific condition, it lacks directional adaptability due to the absence of wind angle modifiers. ASHRAE's model includes generalized effectiveness coefficients for perpendicular and diagonal winds but also fails to respond dynamically to varied wind directions. In contrast, Wang's model incorporates directional modifiers and produced more flexible results that aligned with CFD outputs (RMSE = 0.06). However, it exhibited a steep increase in flowrate with wind speed, highlighting a need for calibration under variable conditions.

There are some issues to increase the accuracy of the SEM model by using Wang's model as a baseline model. According to equation 5, Wang's SEM takes the height of the neutral plane (z_0), pressure coefficients (C_p) and reference wind speed (U_{ref}) as main parameters. Wu et al. (2021) have shown the importance of getting a reasonable neutral plane level for the different shapes of windows. Moreover, it is challenging to apply those models to the whole building since it does not consider the impact of surroundings and the decreased wind speed around openings. The identified problems are described in Fig. 2. The first problem is the adjustment in the level of the neutral plane for the specific window type. Since different wind direction causes the different turbulence patterns, knowing the impact of the window's shape from the various wind incident angle is crucial. The second problem is the lack of understanding of decreased wind speed around the building from the reference location far away from the building. It can be wisely considered and applied the local wind speeds with a simple modifier before calculating Q with the wind velocity value. The last problem is the lack of consideration of the surrounding buildings.

Based on the identified problems above, four different methods were applied to develop Wang's SEM model, which is the most relevant model since it considers the impact of hopper windows on ventilation rate. To calibrate Wang's models, we used outputs of the steady-state CFD simulations [33], the localized wind speed in the urban setting, and the numerical correlations from the literature reviews [35]. Fig. 3 delineates the conceptual understanding of the corrected factor for the model and

Table 2
The calculated Q for different wind incident angles and speeds (Wu et al, 2021).

Case	Window Opening Angle	Wind Incident Angle	Wind Speed(m/s)	ASHRAE (1987) (m^3/s)	Warren (1977) (m^3/s)	Wang (2015) (m^3/s)	Wu et al. (2021) (m^3/s)
1_1	30	0	3.5	1.478	0.045	0.128	0.049
1_2	30	45	3.5	0.568	–	0.095	0.044
1_3	30	90	3.5	0.568	–	0.091	0.066
1_4	30	135	3.5	0.568	–	0.101	0.01
1_5	30	180	3.5	0.568	–	0.094	0.082
2_1	30	0	3	0.975	0.038	0.109	0.04
2_3	30	0	5.4	1.755	0.069	0.197	0.077
2_4	30	0	9	2.925	0.116	0.329	0.136
2_5	30	0	12	3.9	0.155	0.439	0.189

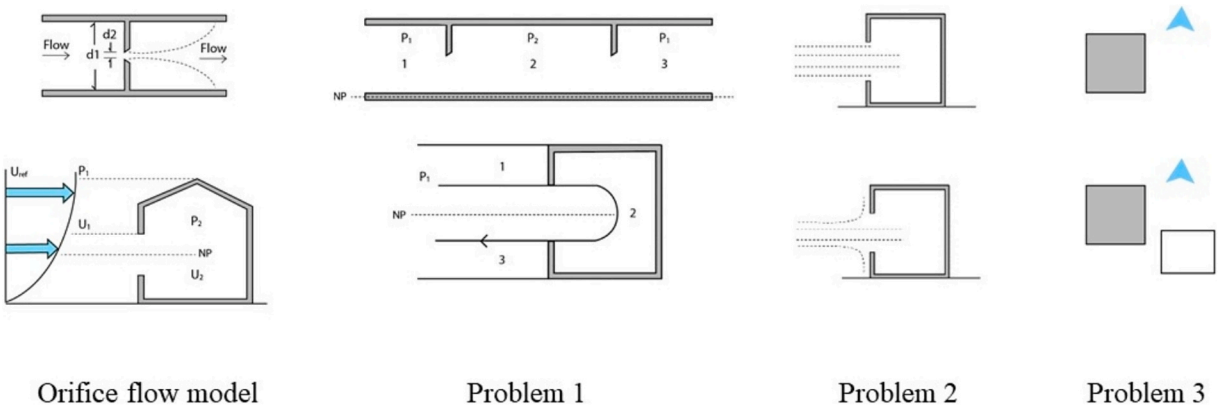


Fig. 2. Identified problems to solve issues on modeling a semi-empirical model.

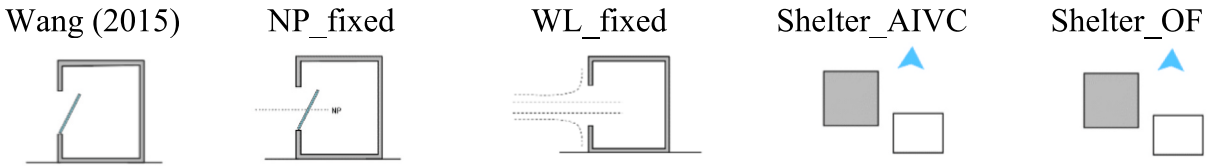


Fig. 3. The diagram of corrected models and name for the SEM model development.

the names for analyzing the calculated Q from SEM. Firstly, we named the calibrated model for the neutral plane [33] to NP_fixed and called the added local wind speed modifier [7] to the NP_fixed model as WL_fixed. Another important factor we should consider when we use the SEM model is the shelter effect from the buildings surrounded. To take the shelter effect into considerations, we replaced C_p from Wang's formula with CFD experiment data. Charisi et al. formulated localized C_ps from standard datasets and CFD simulations. We tested two different methods for applying the effect of surrounding buildings: Shelter_AIVC, which uses the datasets from the Air Infiltration and Ventilation Centre (AIVC, Malcolm Orme, Martin Liddament (1998)) and Shelter_OF, which takes the datasets from the OpenFOAM CFD simulations [36,37].

Fig. 4 shows the results of SEMs for different wind incident angles (Fig. 4 on the left) and the different wind speeds (Fig. 4 on the right). Wang's model shows an average flowrate of 0.102 m^3/s for different wind incident angles and 0.274 m^3/s for the different wind speeds.

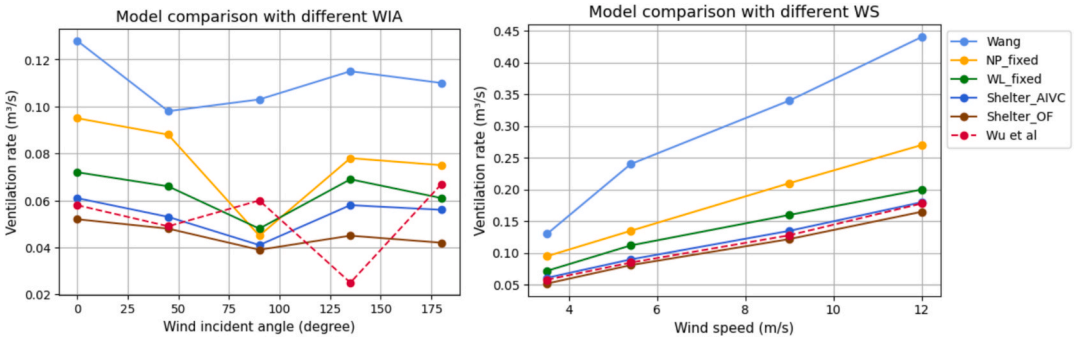


Fig. 4. Flowrate with different wind incident angles (left) and varying wind speeds (right).

Compared to the CFD simulations results, the average flowrate from Wang's model almost doubled. This is because we didn't consider the proposed issues in Fig. 3. Compared to Wang's model, corrected models present a more comparable Q from the CFD. The yellow line graph shows the most significant drop from the baseline Wang's model by replacing neutral planes for each different angle for wind direction and speed [33]. Another significant impact for the SEM calibration for the lab is replacing U with the localized wind speed (Greenline in Fig. 4, WL_fixed), which is usually smaller than the wind from the free stream. When we corrected the local wind value from the reference wind speed at 3.5 m/s, the root mean squared error (RMSE) dropped about half for the different wind incident angles and decreased significantly for the wind speed. It has been found that by replacing the localizing wind speed and the level of the neutral plane, the error between SEM and CFD has decreased by a significant amount. Last modifications were done with corrected C_ps simulating the partially sheltered buildings compared to the building in the open location. These options result in the best-fitted results to CFD, especially for the different wind magnitudes. However, the CFD results were more deviated from other models for the wind direction between 90 degrees to 180 degrees; this is because those directions are parallel to the facades or behind the target facades. However, except for those directions, corrected models predict flowrate very well for the varied wind speeds and wind directions normal to the opening.

With all parameters fixed, including the neutral plane, the wind local, and the shelter effect, the total RMSE errors went down to 0.033 and 0.005 for the different wind directions and speeds. These correction factors especially working well for the different wind magnitudes yielding the lowest errors among all tests. Fig. 4 highlights the performance gains achieved through the corrections, particularly under varying wind magnitudes, confirming the necessity of contextual calibration for accurate flow rate prediction.

To enhance the evaluation of classical semi-empirical models, we compared ASHRAE (1987), Warren (1977), and Wang (2015) against CFD results using a shoebox test building. While ASHRAE and Warren offer simplicity, they lack adaptability to wind direction, often resulting in over- or underprediction. Wang's model performs better due to its inclusion of wind direction and window geometry but still overestimates flowrates at higher wind speeds. To address these limitations, we applied four calibration strategies to Wang's model: neutral plane adjustment (NP_fixed), local wind speed correction (WL_fixed), and two shelter effect models (Shelter_AIVC, Shelter_OF). These corrections significantly reduced RMSE, with WL_fixed and Shelter_OF yielding the most accurate predictions. This highlights the importance of localized calibration for accurate natural ventilation modeling. These findings underscore that, while traditional SEMs are valuable for early-stage design, they must be critically evaluated and contextually calibrated to provide meaningful predictions in real-world applications. Our enhanced framework demonstrates a robust methodology for adapting and refining existing SEMs through empirical validation and site-specific data integration, thereby enabling more accurate and scalable applications in natural ventilation modeling. Given enhanced model performance through contextual calibration, the following section applies these corrections to a controlled laboratory setting. Since pressure coefficients (Cp) are a critical input for flowrate estimation, it is essential to validate their applicability under real experimental conditions and to develop robust data-driven models for Cp prediction.

3. Methodology for the ML-guided SEM model

Semi-empirical models (SEMs) rely on various inputs such as window opening angles, effective opening areas, neutral plane height, wind conditions, and façade pressure distributions. While wind speed and direction can typically be obtained from local weather stations, other parameters—Cp values—are often derived from standardized empirical datasets, which may not reflect site-specific turbulence or the influence of surrounding buildings. To overcome these limitations, this study proposes a machine learning-guided SEM (ML-SEM) that incorporates real-time sensor data to predict localized pressure coefficients (Cp).

The methodology consists of three integrated components: (1) constructing a single-sided ventilation model adapted to the laboratory setting; (2) training a machine learning model to predict Cp using sensor-derived features; and (3) developing a workflow that links sensor inputs, ML outputs, and the SEM to improve the accuracy and contextual relevance of airflow estimation. By embedding a data-driven Cp prediction module, the ML-SEM reduces reliance on generalized assumptions and enhances model adaptability to real-world conditions.

3.1. Experiments for machine learning training and validation

3.1.1. Experimental building and measurement

The space utilized for single-sided empirical model development was a lab in HouseZero (Fig. 5a). HouseZero is located in Cambridge, Massachusetts and has hundreds of sensors utilized to run the building and conduct experimentation [38]. The third floor of HouseZero contains a controllable experimental area known as the LiveLab, which is oriented toward the south. This lab is integrated with building control systems that automatically operate the hopper windows of HouseZero (Fig. 5b).

The LiveLab is equipped with several types of sensors that provide data to train and validate the ML algorithm developed for the data driven model (Fig. 6a). Sensors installed in HouseZero include temperature sensors, pressure transducers, anemometers, occupancy sensors and a weather station. The local weather station is mounted (See Fig. 6b) at the HouseZero rooftop. The weather information includes dry-bulb temperature, relative humidity, wind direction, wind speed and solar radiation. Eleven pressure and temperature sensors are installed on all facades. Fig. 6s shows the location of facade sensors and Table 3 describes important sensor information: accuracy and precision. The weather station measures wind speed with an accuracy of ± 0.3 m/s for speeds below 3 m/s and ± 1 % for speeds above 3 m/s, within a range of 0.3 to 75 m/s. Wind direction is measured with an accuracy of ± 2° across the full 0° to 359° range. Temperature sensors report air temperature with high precision: ±0.25 °C from −40 °C to 0 °C, ±0.20 °C from 0 °C to 70 °C, and ± 0.25 °C from 70 °C to 75 °C, covering an operational range from −30 °C to 55 °C. The barometric pressure sensors operate within a range of 60 to 110 kPa and have an accuracy of ± 0.03 % FS, where "FS" refers to the full-scale measurement range, corresponding to a maximum possible error of ± 0.015 kPa.

The experimental period spanned from June 1, 2022, to May 31, 2024. Temperature, pressure, and weather data were continuously collected throughout this period. For this study, pressure data were obtained from the sensor installed on the south façade of the LivingLab (see Fig. 6c), recorded at a 5-minute interval over the full duration of the experiment.

3.1.2. Data from the weather station and the pressure sensor

Typically, the weather data from the nearest airport is used for the performance simulation. Sometimes weather data from the airport cannot represent the localized weather condition. For example, the average dry bulb temperature from the HouseZero, which is located in the middle of the urban center, is more than 4 °C higher than the airport [34]. Fig. 7 illustrates the difference between wind information from the nearest airport and localized weather station.

For this study, the data collected from the HouseZero weather station was merged into the datasets from the nearest airport − Boston Logan International Airport. The utilized data from the airport was mainly the atmospheric pressure data since it is not located in the dense urban areas. The rest of the datasets including temperature, humidity, wind direction and speed was directly adopted from the local weather station. As shown in Fig. 8, the wind speed was widely distributed from 0 m/s to 4 m/s at HouseZero, but the wind direction was concentrated between 200 degrees and 315 degrees. Both wind speeds and directions were

Fig. 5. (a) HouseZero, (b) Hopper windows.

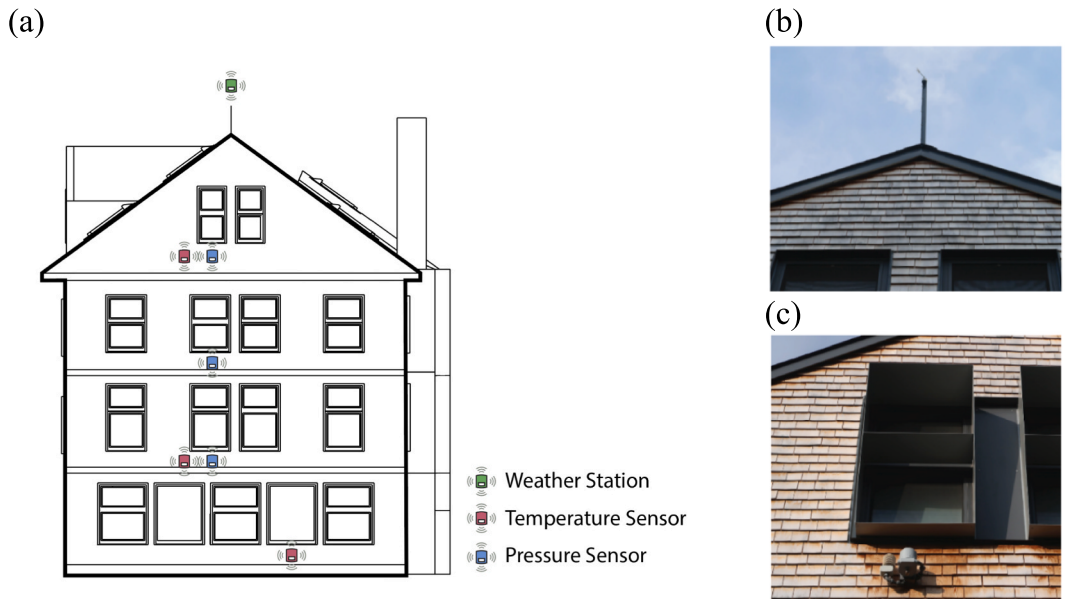


Fig. 6. (a) The diagram of sensors on HouseZero's south facades (b) Weather station on the roof top (c) The temperature and pressure sensor on the south façade.

Table 3
The details of sensors installed.

Name	Objective	Accuracy	Range
Weather station	Wind speed	±0.3 m/s below 3 m/s ±1% over 3 m/s	0.3 to 75 m/s
	Wind direction	±2°	0° to 359°
Temperature sensor	Air temperature	±0.25 °C from −40° to 0 °C ± 0.20 °C from 0° to 70 °C ±0.25 °C from 70° to 75 °C	−30 to 55 °C
Pressure sensor	Barometric pressure	±0.03 % FS	60 to 110 kPa

used to predict C_ps with pressure data from the airport and a pressure sensor on the south façade. The C_p values are then fed into the fixed Wang's model with fixed wind speed and neutral plane for the lab. The level of neutral plane was calculated by using CFD with the methods described in the section 2.

3.2. *ML-based C_p model development for semi-empirical model*

To address the limitations of conventional pressure coefficient (Cp) estimation methods in semi-empirical models (SEMs), this study incorporates a ML approach to dynamically predict Cp values using real-time environmental data. Specifically, a Random Forest regressor was employed as the primary ML model, leveraging sensor inputs from the HouseZero building and a localized weather station. This data-driven Cp model replaces the conventional Cp component derived from the literature, thereby improving the model's contextual responsiveness and predictive accuracy.

Fig. 9 illustrates the workflow for integrating ML model into the SEM model for predicting natural ventilation flowrates. The process begins with the collection of sensor data, including weather variables (e.g., wind speed and direction) and pressure measurements. Weather data is used to construct input features, while pressure data is processed to calculate pressure coefficients (Cp), which serve as labels for supervised learning. These data are then used to train multiple ML models through a

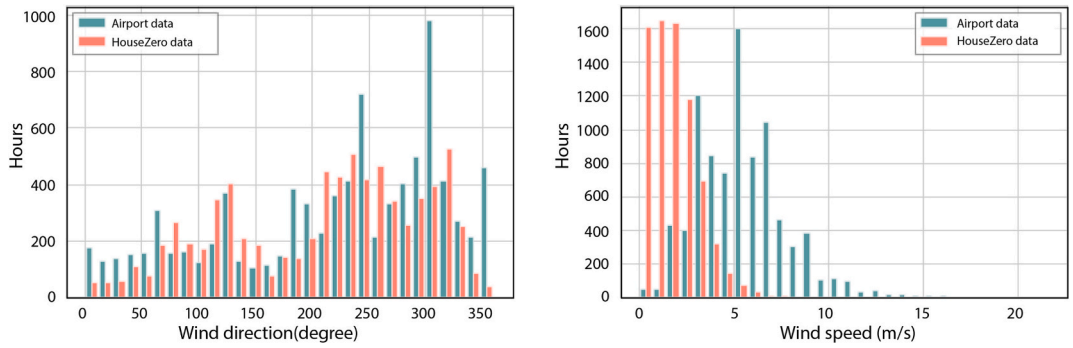


Fig. 7. The histogram of wind direction and wind speed from the different weather station.

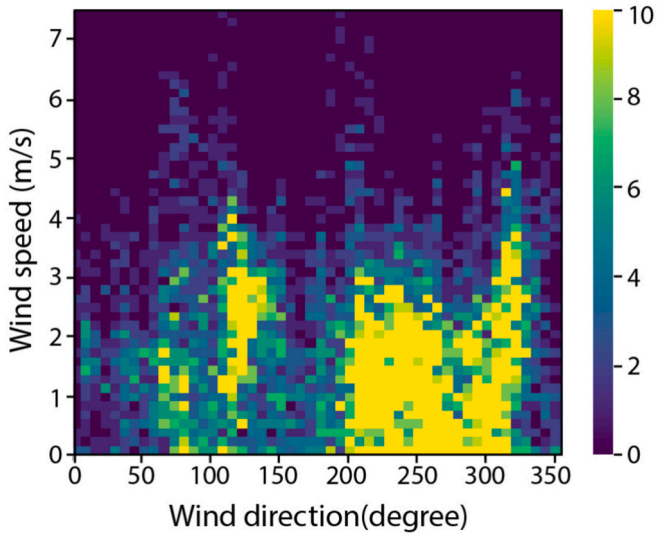


Fig. 8. The heatmap of wind speed and wind direction from the collected weather data.

process of model comparison, feature selection, and hyperparameter tuning. Once the optimal model—such as a random forest regressor—is selected, it is applied to new sensor data to predict Cp values in real-time. These predicted Cp values are then passed to the SEM, which computes the resulting airflow rate through the building's openings. In the following section, we describe the detailed procedure for developing and validating the ML model, including feature selection, algorithm tuning, and performance evaluation based on changes in wind speed and direction.

3.2.1. Machine learning (ML) algorithm

1. Support vector machines (SVM)

SVMs are effective in high-dimensional spaces and particularly useful when the relationship between input features and the target variable is non-linear. The use of kernel functions allows SVMs to project data into higher-dimensional spaces, making it possible to delineate complex patterns that linear models cannot capture. This capability is crucial for accurately modeling the intricate interactions between environmental variables and pressure coefficients.

2. Decision trees

Decision trees provide a straightforward approach to modeling non-linear relationships by recursively splitting the data based on feature values. Each split represents a decision rule that leads to a clear and interpretable model structure. The flexibility of decision trees in handling various types of data and their ability to model non-linear interactions without requiring any transformations of the input variables to make them a suitable choice for our task.

3. Random Forest

Random forests enhance the predictive power of decision trees by constructing a multitude of trees and averaging their predictions. This ensemble approach improves the model's accuracy and reduces the risk of overfitting. Random forests can effectively capture complex, non-linear relationships in the data by leveraging the diversity of the individual trees, making them robust against the variability in environmental conditions that influence pressure coefficients. Fig. 10 shows the structure of the model, and it contains parallelly running decision trees with no interaction amongst them.

3.2.2. ML model development and fine-tuning

We compared the general performance of the models described in the previous section using the coefficient of determination (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE) on validation sets to evaluate their effectiveness in predicting pressure coefficients.

SVM demonstrated limited effectiveness in capturing the non-linear relationships between environmental inputs and pressure coefficients, with an R^2 of 0.421, MAE of 0.0309, and MSE of 0.0008. Decision Trees showed slight improvement with an R^2 of 0.487, MAE of 0.0182, and MSE of 0.0006. Among the evaluated models, Random Forests delivered the best performance, achieving the highest R^2 value of 0.619, the lowest MAE of 0.0159, and the lowest MSE of 0.0005 (see Table 4). The ensemble nature of Random Forests enables robust learning from complex patterns while reducing overfitting risks, making them the most suitable approach for predicting localized pressure coefficients (C_p) from environmental sensor data.

Therefore, Random Forest regressor was selected as the machine learning model to predict localized C_p. A total of 17,520 hourly data points representing a full year were used. The dataset was randomly split into 80 % for training and 20 % for validation, with an entirely separate dataset used for final testing. Fig. 11 presents scatter plots that compare predicted and measured C_p values across three machine learning models—SVM, Decision Tree, and Random Forest—applied to both training and test datasets. The top row shows results for the training set, while the bottom row corresponds to the test set.

To determine the most effective input variables for predicting C_p, a feature selection analysis was conducted using the trained Random Forest regressor. As shown in Fig. 12, wind direction emerged as the most influential predictor, followed by air temperature and relative humidity. These rankings were derived from the model's internal feature importance metric, which evaluates each variable's contribution to reducing prediction error.

The inclusion of temperature and relative humidity as predictors is theoretically justified by the general energy equation, which establishes relationships between sensor-measured environmental parameters and surface-level pressure coefficients at natural ventilation openings. These environmental variables effectively capture the thermal and moisture-related interaction that influence pressure distributions on building surfaces.

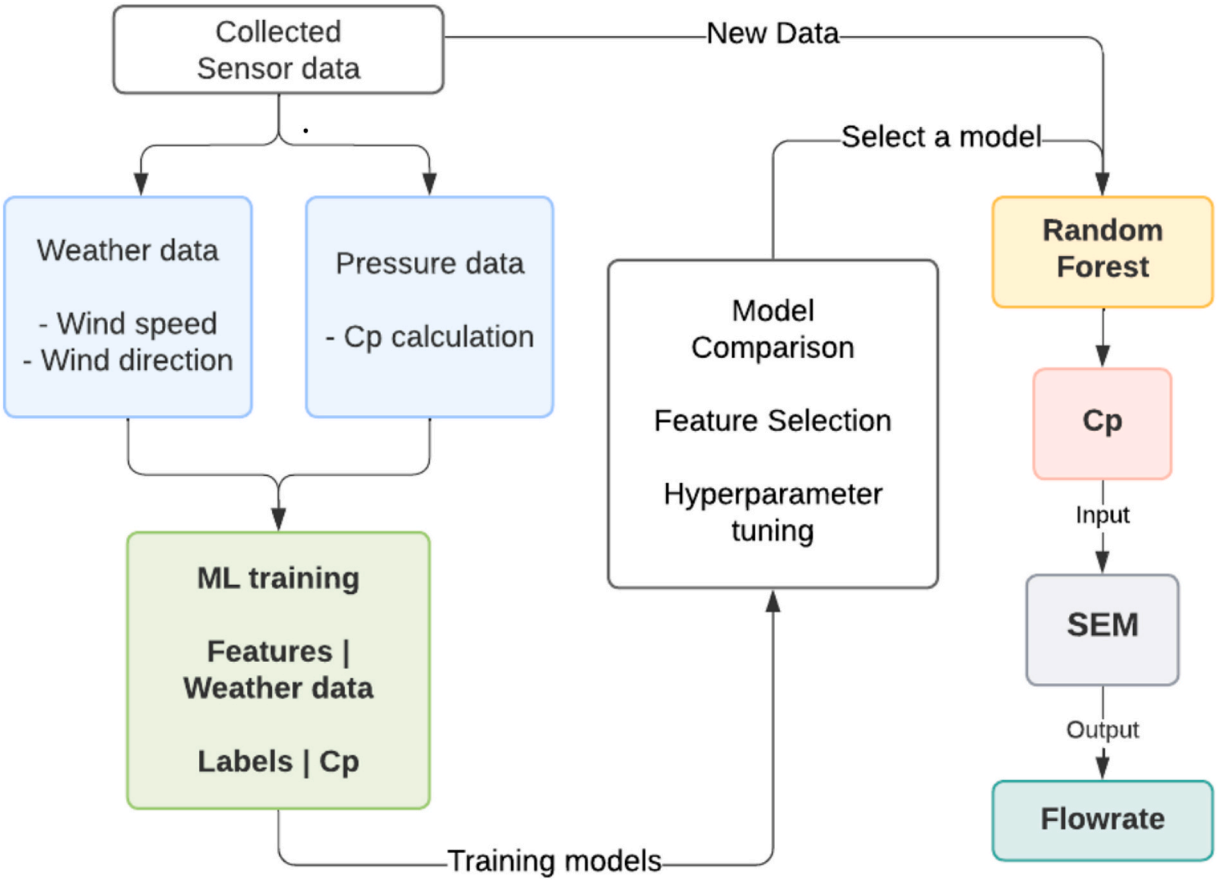


Fig. 9. Flowchart of ML-based flowrate calculation.

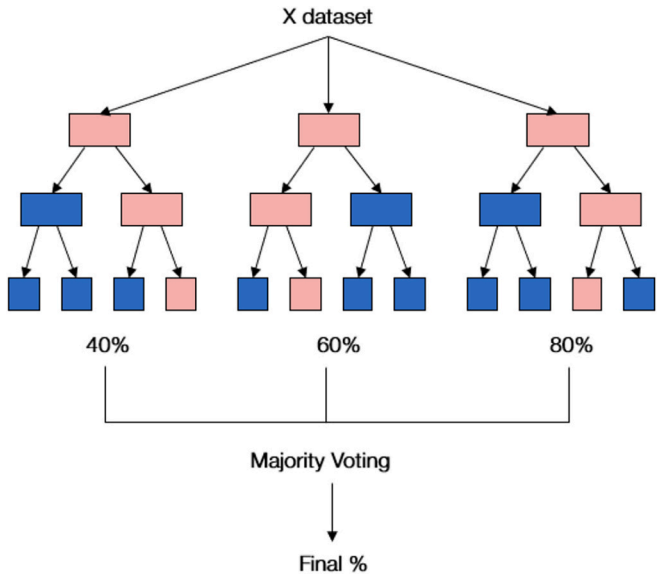


Fig. 10. Random forest regressor diagram.

Table 4
Errors on validation sets for different ml models.

	SVM	Decision Tree	Random Forest
R2	0.421	0.487	0.619
MAE	0.0309	0.0182	0.0159
MSE	0.0008	0.0006	0.0005

$$P_m + \frac{\rho_m v_m^2}{2} + \rho_m g z_m = P_s + \rho_s g Z_s + loss \tag{7}$$

where subscript m means the parameter at the measurement location, s means the opening surface location. Both temperature and relative humidty affects the local air density and leads to pressure differences.

This relationship implies that temperature and humidity, which directly influence air density ρ, are key drivers of pressure differences across the building envelope. Variations in thermal and moisture content between indoor and outdoor air masses can lead to localized buoyancy and dynamic effects that modulate pressure coefficients at façade openings. Therefore, the prominence of temperature and humidity variables in the feature ranking not only validates their empirical relevance but also aligns with theoretical expectations. Their inclusion enhances the model's ability to infer C_p values under diverse environmental conditions, improving its robustness for real-time application in data-driven SEMs.

Motivated by these findings, we designed a comparative experiment to evaluate the predictive performance of different feature configurations. Specifically, we compared three representative input scenarios: (1) using wind direction only, (2) using the top three features identified via feature importance (wind direction, air temperature, and relative humidity), and (3) using the full set of available features.

Among these, the top three feature set provided the best balance between model parsimony and predictive power. It significantly improved performance over the wind direction–only model, increasing the R^2 from 0.534 to 0.640 and reducing the mean squared error (MSE) by over 10 %. In contrast, the performance gain from adding all remaining features to the top three was marginal, yielding less than a 1 % decrease in MSE and nearly identical mean absolute error (MAE) values. As illustrated in Fig. 13, the predicted C_p values from the top

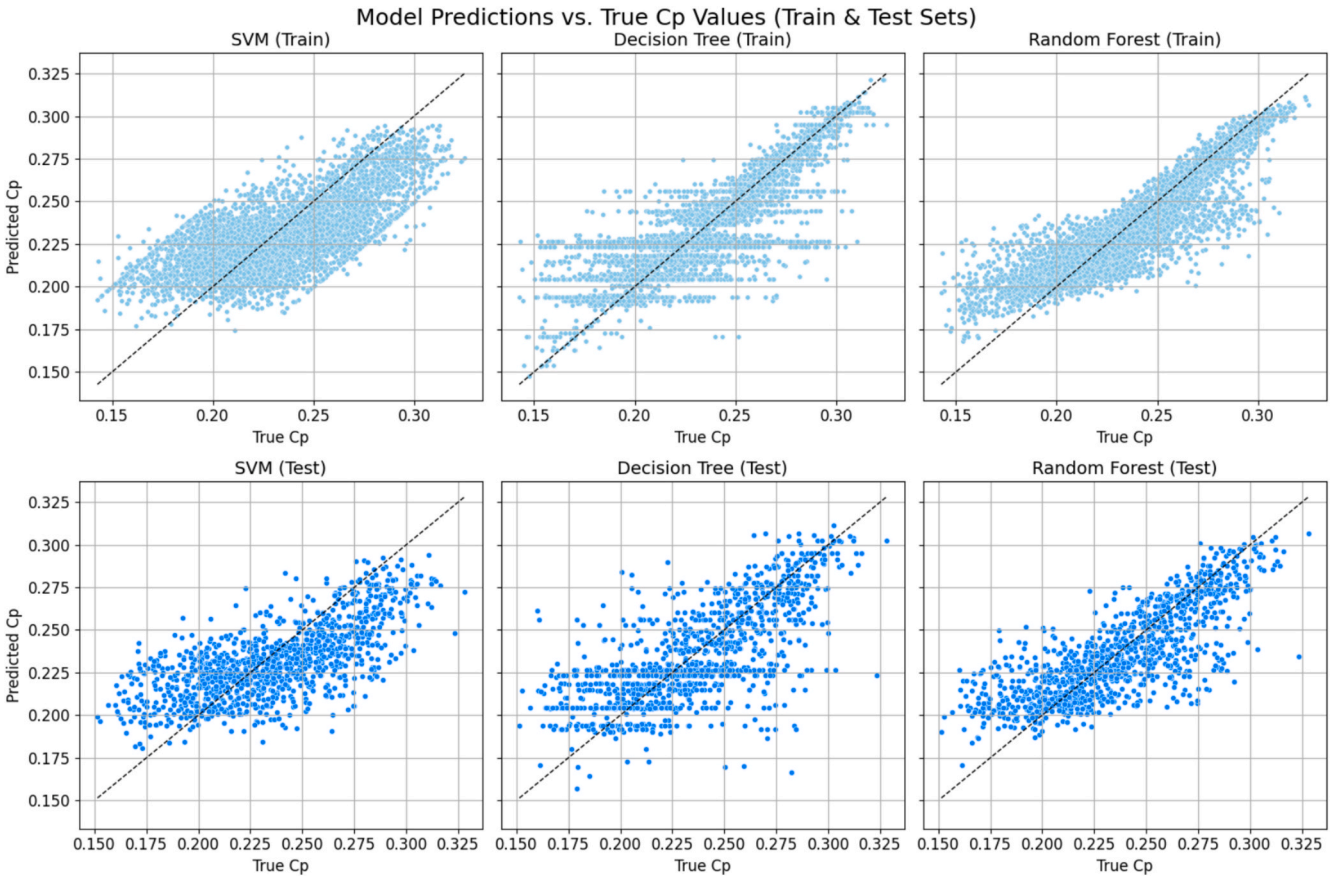


Fig. 11. Scatter plots of predicted vs. true C_p values for three machine learning models (SVM, decision tree, random forest) on both training and test sets.

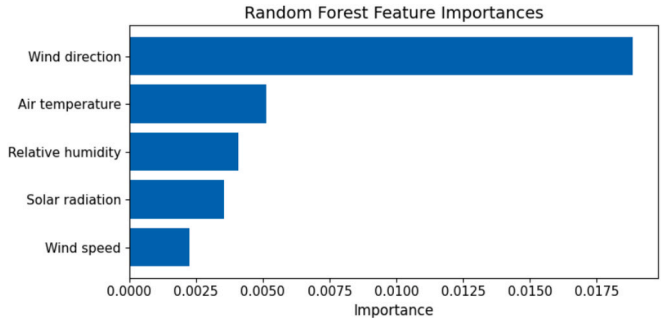


Fig. 12. The feature importance and score chart.

three feature model closely aligned with those from the full-feature model, whereas the wind direction–only model exhibited larger deviations from ground truth.

After selecting wind direction, wind speed, air temperature, and relative humidity as the core input features, we further investigated whether adding additional variables could enhance model performance. We tested three types of supplementary inputs: time-related features (such as sine and cosine transformations of hour and month), interaction terms (e.g., temperature × solar radiation, humidity × wind), and lagged C_p values from previous time steps (see Table 5). The inclusion of time-related features improved the model's R2 from 0.640 to 0.675, while reducing the MAE and MSE to 0.029 and 0.006, respectively. This suggests that capturing daily and seasonal patterns contributes to more accurate C_p predictions. Similarly, adding interaction terms yielded comparable results, with R^2 at 0.693 and a slight reduction in error

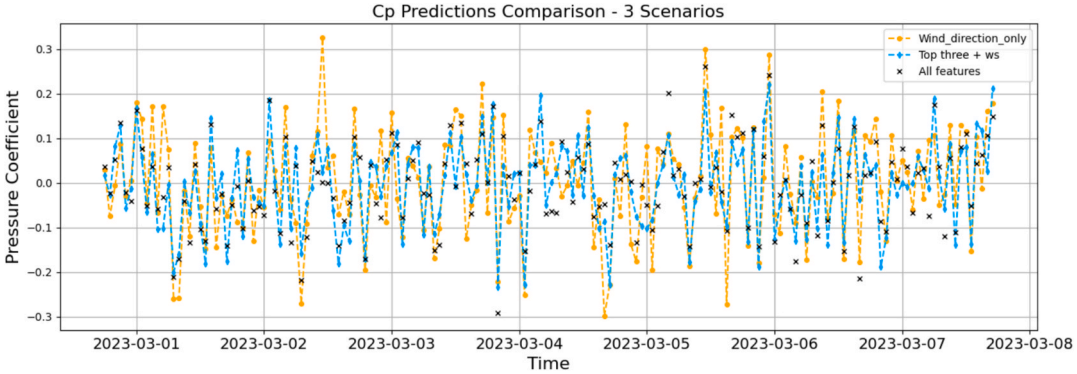


Fig. 13. The predicted C_ps between different inputs features.

Table 5
Error on validation sets for different input feature configurations.

	WindDir_only	Top three	All features	Time info	Interaction	Lag C_p
R2	0.534	0.640	0.615	0.675	0.698	0.922
MAE	0.046	0.035	0.033	0.029	0.013	0.017
MSE	0.009	0.007	0.007	0.006	0.004	0.001

metrics, highlighting the benefits of modeling non-linear relationships between environmental variables. The most notable improvement came from incorporating lagged C_p values, which significantly boosted the model's R^2 to 0.922 and reduced the MAE and MSE to 0.017 and 0.001, respectively. This indicates a strong temporal dependency in C_p behavior, and that leveraging historical C_p data substantially improves predictive accuracy. These results confirm that while the top three features provide a solid baseline, adding time-aware and history-based features can further refine the model, especially in contexts where pressure data is available.

To optimize hyperparameters, the grid search cross-validation methods were used. Table 6 describes the search space used and the best parameters for the final SEM model. This approach ensures that the model is fine-tuned to achieve the highest possible accuracy by systematically exploring different combinations of hyperparameters. The selected parameters, such as the number of estimators and maximum depth, reflect a balance between model complexity and performance. By incorporating these optimized hyperparameters, the random forest regressor can more accurately capture the relationships within the data, thereby improving the prediction of C_p values.

The fine-tuned model was implemented in SEM as a part of the model to calculate C_p. The fine-tuned model shows the mean absolute error at 0.021 and the mean squared error at 0.002 on the final testsets. Test sets are the data that are not included in both the training and validation process.

Fig. 14 presents the one-week time-series comparison of predicted and actual C_p values using both the baseline and the fine-tuned Random Forest models. In each plot, the solid line represents the model predictions, while the shaded area denotes the $\pm$ 1 standard deviation confidence interval estimated via bootstrap resampling. The black markers indicate the measured values. Notably, the final model (bottom figure) shows improved alignment with observed C_p fluctuations compared to the baseline (top figure), capturing short-term temporal variations more accurately. Additionally, approximately 75 % of the true values lie within the model's confidence bounds, highlighting the enhanced predictive stability and generalization of the fine-tuned model over the baseline model.

4. Results and discussion

It is crucial to tune the parameters which represents a location-specific case. There have been many ways to calibrate SEM including CFD, experimental data, and surrogate models. We utilized measured data to formulate surrogate model as a part of the SEM. Fig. 15 illustrates the processed data with credible intervals to compare the empirical C_p models for the HouseZero LiveLab compared to the traditional C_p models. In literature, the wind from perpendicular to the façade has the most potent impact on natural ventilation. The empirical formula from ASHRAE and Wang also shows this idea well by the curve on the peak at the wind incident angle at 180 degrees. However, our data tells a different story. For the wind incident angle at 180 degrees, the ML-based C_p model shows the lowest C_p compared to slightly tilted angles. This might be due to the context building near the south façade blocks direct wind from the south and dissipates the southern wind towards the south-faced façade.

It is also found that the wind parallel to the façade shows a significant spark in the predicted C_p values. With those findings, we can define the unknown factors around buildings and the local turbulence caused by contextual buildings and conditions. This will directly affect the flow-rate on the target room using the proposed empirical formula. We used the empirical formula in section 2, and modify the C_p values accordingly: ASHRAE model, Wang's model, and data-driven model in the following orders.

Due to the development of simplified tools and advances in CFDs, more radical models and methods have been proposed for natural ventilation models. However, using a CFD approach is relatively complex for accurate modeling of the airflow around the buildings and considering the urbanized local context in the model [39]. As a result, the SEM models have been concurrently developed and validated using field measurement data and CFDs data. Developing fully matured SEM utilizing on-site measurement takes a fair amount of time and requires manifold sensors to point and control the environment. To avoid this, calibrated SEM was alternatively widely used in many cases, such as model predictive controls, window controls, and indoor air quality monitoring. To increase SEM's accuracy, we proposed a data-driven component as a part of SEM to represent a site-specific condition well. The results from the previous section showed the various considerations around buildings by calculating the pressure coefficient on the façade. In this paper, a random forest regressor with weather related input features was created and trained to obtain localized C_p values. Data used were collected from the sensors installed in HouseZero. We used the trained C_p model to calculate flowrate for the future use of the building controls and other experiments. The trained model yields the lowest MSE for the test sets at 0.008 within 75 % of the credible boundary on the testsets.

While the model performed well, it is important to acknowledge the limitations in temporal resolution. High-frequency data is typically more suitable for capturing detailed turbulence effects. However, due to equipment limitations and data storage constraints, this study utilized 5-minute interval data. This resolution is sufficient for capturing broader patterns and average behavior in natural ventilation and is consistent with the time resolutions commonly used in empirical research.

Fig. 16 shows the calculated flowrate from the same SEM model with different C_ps values. It illustrates the effect of C_ps calculating flowrate is relatively large when wind incident angle changes. Therefore, it is difficult to conclude that the generalizability of SEM is flexible when it comes to apply assumptions on predictions. Therefore, it shows the importance of implementing well-made assumptions to utilize existing SEM for the different projects. In the case of this research, the ML-guided C_p model reduces the unpredictability of the wind pressure around buildings by using the measured data located in building's façades. The proposed methods can be extended to any other projects where sensors' datasets are available such as temperature, relative humidity, and wind direction. Wind speed can be added or omitted based on the available datasets.

The results indicate that incorporating ML with semi-empirical models (SEMs) significantly enhances the prediction accuracy of natural ventilation (NV) rates in buildings. Using real-time data from the HouseZero provided a novel approach to use the sensor's data for modelling NV, demonstrating that ML models can effectively capture local environmental conditions and adjust ventilation predictions

Table 6
Radomized grid search parameters for the proposed regressor.

Parameter	Grid Search Space	Best Value
bootstrap	[True, False]	True
max_depth	[1, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, None]	30
max_features	['auto', 'sqrt']	'sqrt'
min_samples_leaf	[1,2,4]	2
min_samples_split	[2,5,10]	10
n_estimators	[200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]	1000

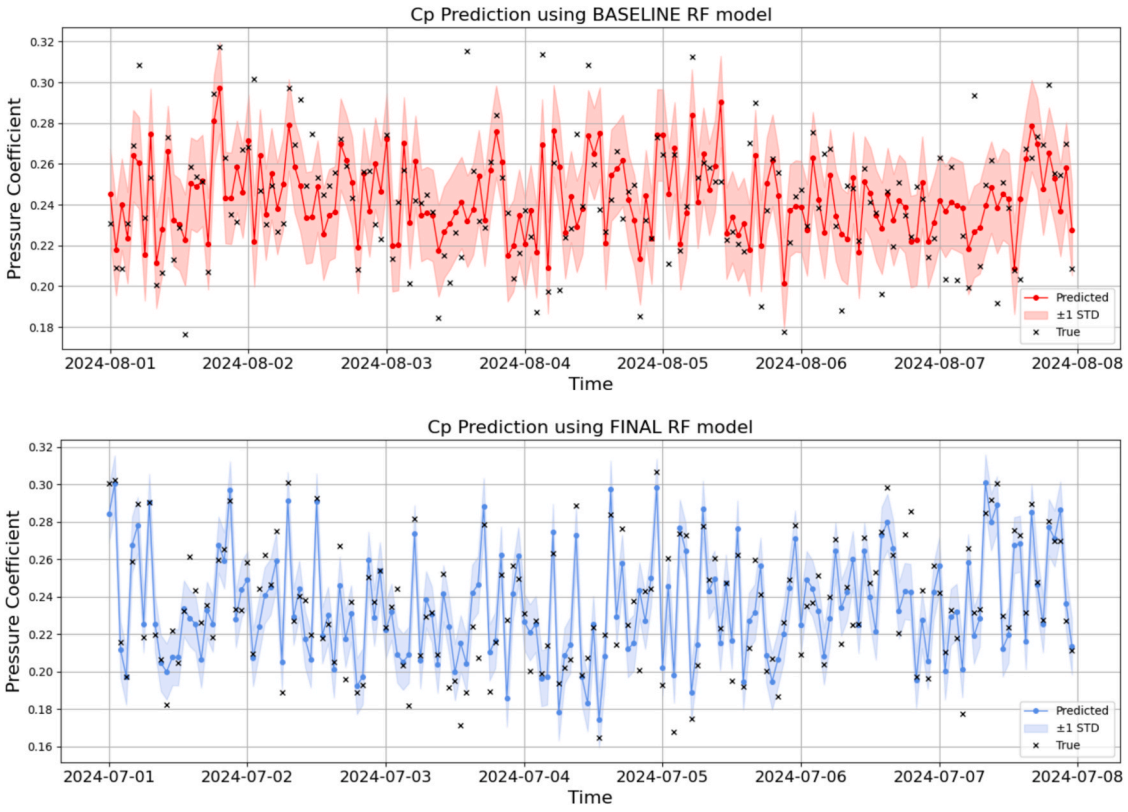


Fig. 14. The predicted C_ps on test sets compared to the actual values with final models.

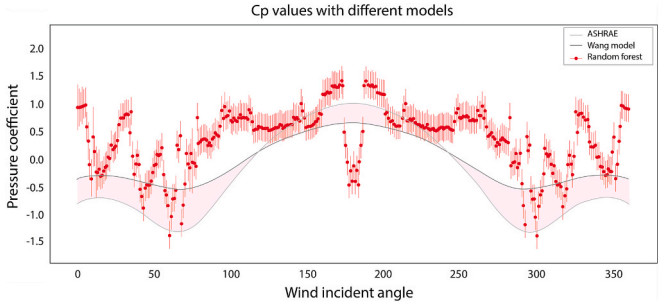


Fig. 15. The predicted C_ps from the ML model, Wang's model and ASHRAE model.

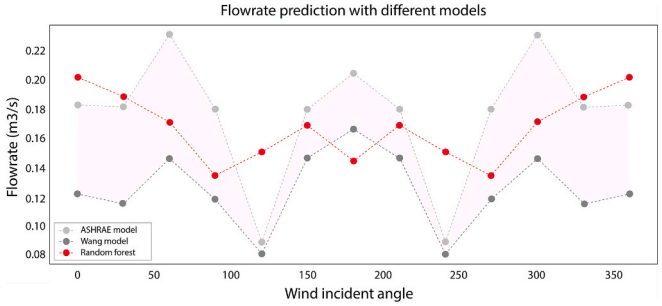


Fig. 16. The flowrate prediction with different wind incident angles.

accordingly. Our proposed methodology addresses several limitations of traditional SEMs, particularly the reliance on generalized C_p values that do not account for specific building contexts. By integrating localized sensor data and employing an ML model to predict C_p values, our approach mitigates these inaccuracies, providing more precise and reliable ventilation rates.

The results from the random forest regressor training on localized C_ps reveal that wind direction is the most influential factor in predicting C_p values, highlighting its significant impact on airflow patterns around building facades. Following wind direction, temperature and relative humidity were also critical for prediction accuracy due to their effects on air density, buoyancy forces, and thermal properties. Wind speed, while the least important, still contributed to the prediction of these models. Our analysis of five scenarios, ranging from using only wind direction to all available features, showed that including temperature, relative humidity, wind direction, and wind speed provided a comprehensive and accurate model, with negligible MSE differences between using all features versus excluding wind speed. A significant performance gap was noted between models using only wind direction and those using the top three features, indicating that a multi-variable approach reduces error by over 10 %. Consequently, our model prioritizes temperature, relative humidity, wind direction, and wind speed as input parameters. However, in the absence of wind speed sensors, the model remains robust with just wind direction, temperature, and relative humidity. This flexibility ensures wide applicability and reliable C_p predictions across various scenarios, emphasizing the importance of a comprehensive feature set for accuracy and adaptability in different real-world applications.

Despite the promising results, our study has several limitations. The accuracy of the ML model heavily relies on the quality and quantity of the sensor data, and any inaccuracies or gaps can affect the model's predictions. The specific environmental and architectural context of the HouseZero may also limit the generalizability of our findings to other buildings with different configurations and climates. However, the modelling techniques are reproduceable that other researchers might benefits from predicting C_p. Additionally, while our model significantly improves C_p predictions using localized data, C_p values can still be influenced by unmodeled external factors such as unexpected weather

changes or building use patterns, thus future fine-tuning with proper data augmentation would be recommended.

To build on the findings of this study, future research could explore several avenues. Conducting validation studies across different seasons and various building types would provide a more comprehensive understanding of the model's robustness and generalizability. Increasing the resolution of sensor data and incorporating additional environmental variables such as localized wind speed and direction could further enhance the model's accuracy. Applying the ML-guided SEM model to different buildings with diverse architectural designs and in various climatic regions would help generalize the findings and refine the model for broader application. Integrating the ML-guided SEM with automated building management systems could create adaptive ventilation strategies that respond dynamically to changing environmental conditions and occupant needs.

5. Conclusion

This paper developed a machine learning-guided semi-empirical model (ML-SEM) and proposed a training workflow to generalize the approach for future applications. By integrating C_p estimation based on extensive sensor data from an operational building, the model improves predictive accuracy for wind-driven single-sided natural ventilation. The ML model effectively captures local conditions without relying solely on CFD simulations or generalized empirical equations.

A random forest model was trained using sensor data—specifically wind direction, temperature, and relative humidity—as these variables were found to be the primary drivers of localized C_p variations. The predicted C_p values were directly embedded into a validated SEM to calculate flowrates through a hopper window, demonstrating the practicality and robustness of the proposed framework.

Moreover, the integration of real-time sensor data allows the model to implicitly reflect the effects of surrounding buildings, upstream turbulence, and microclimatic variations, which are often difficult to model explicitly. This data-driven workflow offers a scalable and adaptive methodology for incorporating sensor-derived pressure data into SEMs, thereby reducing uncertainty and improving reliability in natural ventilation assessments.

The broader implication of this work lies in minimizing the reliance on generalized assumptions typically used in empirical models. Future research may focus on improving temporal resolution of the dataset to support the development of more advanced models tailored to site-specific conditions. Additionally, validating the predicted flowrates against direct airflow measurements in the HouseZero experimental lab could further enhance the reliability and applicability of the proposed ML-SEM framework.

CRediT authorship contribution statement

Jung Min Han: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Wentao Wu:** Writing – original draft, Supervision, Resources, Conceptualization. **Ali Malkawi:** Writing – review & editing, Resources, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgement

This research was supported by the Yonsei University Research Fund of 2024-22-0035.

Data availability

Data will be made available on request.

References

[1] I. Loche, F. Bre, J.M. Gimenez, R. Loonen, L.O. Neves, Balcony design to improve natural ventilation and energy performance in high-rise mixed-mode office buildings, Build. Environ. 258 (2024) 111636. https://doi.org/10.1016/j.buildenv.2024.111636.
[2] Y. Pan, W. Zhong, X. Zheng, H. Xu, T. Zhang, Natural ventilation in vernacular architecture: a systematic review of bioclimatic ventilation design and its performance evaluation, Build. Environ. 253 (2024) 111317. https://doi.org/10.1016/j.buildenv.2024.111317.
[3] Y. Lv, M. Wang, W. Wu, W. Shang, J. Li, H. Yao, G. Quoie, H. Zhang, X. Shen, Experimental Study of Building Natural Ventilation with Pendulum Velocity Anemometer, Build. Environ. Accepted (2024).
[4] Z. Jiang, T. Kobayashi, T. Yamanaka, M. Sandberg, H. Yamasawa, M. Shohei, The similitude of indoor airflow in natural ventilation for a reduced-scale model: investigation of nonisothermal flow fields by RANS simulation, Build. Environ. 262 (2024) 111842. https://doi.org/10.1016/j.buildenv.2024.111842.
[5] Z. Jiang, T. Kobayashi, T. Yamanaka, M. Sandberg, N. Kobayashi, N. Choi, K. Sano, Validity of orifice equation and impact of building parameters on wind-induced natural ventilation rates with minute mean wind pressure difference, Build. Environ. 219 (2022) 109248. https://doi.org/10.1016/j.buildenv.2022.109248.
[6] R. Stasi, F. Ruggiero, U. Berardi, Influence of cross-ventilation cooling potential on thermal comfort in high-rise buildings in a hot and humid climate, Build. Environ. 248 (2024) 111096. https://doi.org/10.1016/j.buildenv.2023.111096.
[7] H. Wang, Q. Chen, A new empirical model for predicting single-sided, wind-driven natural ventilation in buildings, Energ. Build. 54 (2012) 386–394, https://doi.org/10.1016/j.enbuild.2012.07.028.
[8] S. Hou, Y. Liu, Z. Huang, X. Zhao, Experimental study of buoyancy-driven ventilation with a single opening between two adjacent building zones, Build. Environ. 253 (2024) 111331. https://doi.org/10.1016/j.buildenv.2024.111331.
[9] S. Omrani, V. Garcia-Hansen, B. Capra, R. Drogemuller, Natural ventilation in multi-storey buildings: Design process and review of evaluation tools, Build. Environ. 116 (2017) 182–194. https://doi.org/10.1016/j.buildenv.2017.02.012.
[10] P.R. Warren, Ventilation Through Openings on One Wall Only, 1977.
[11] T.S. Larsen, P. Heiselberg, Single-sided natural ventilation driven by wind pressure and temperature difference, Energ. Build. 40 (2008) 1031–1040, https://doi.org/10.1016/j.enbuild.2006.07.012.
[12] P. Heiselberg, M. Sandberg, Evaluation of discharge coefficients for window openings in wind driven natural ventilation, 3315 (2016). https://doi.org/10.1080/14733315.2006.11683723.
[13] H. Wang, P. Karava, Q. Chen, Development of simple semiempirical models for calculating airflow through hopper, awning, and casement windows for single-sided natural ventilation, Energy and Buildings 96 (2015) 373–384.
[14] W. Pan, S. Liu, S. Li, X. Cheng, H. Zhang, Z. Long, T. Zhang, Q. Chen, A model for calculating single-sided natural ventilation rate in an urban residential apartment, Build. Environ. 147 (2019) 372–381. https://doi.org/10.1016/j.buildenv.2018.08.047.
[15] H.L. Gough, J.F. Barlow, Z. Luo, M.F. King, C.H. Halios, C.S.B. Grimmond, Evaluating single-sided natural ventilation models against full-scale idealised measurements: impact of wind direction and turbulence, Build. Environ. 170 (2020) 106556, https://doi.org/10.1016/j.buildenv.2019.106556.
[16] R. Vasaturo, T. van Hooff, S. Gillmeier, B. Blocken, P.J.V. van Wesemael, On the effect of pressure coefficient source on the energy demand of an isolated cross-ventilated building, Build. Environ. 255 (2024) 111436. https://doi.org/10.1016/j.buildenv.2024.111436.
[17] S. De Wit, G. Augenbroe, Analysis of uncertainty in building design evaluations and its implications, Energ. Build. 34 (2002) 951–958, https://doi.org/10.1016/S0378-7788(02)00070-1.
[18] M.S. de Wit, Uncertainty in Predictions of Thermal Comfort in Buildings, PhD Thesis. (2001) 215.
[19] T. Kikuchi, K. Ohtake, Y. Takahashi, H. Watanabe, K. Hidari, S. Tanabe, T. Nobe, Comparison of wind pressure coefficients between wind tunnel experiments and full-scale measurements using operational data from an urban high-rise building, Build. Environ. 252 (2024) 111244. https://doi.org/10.1016/j.buildenv.2024.111244.
[20] X. Xie, Z. Luo, S. Grimmond, L. Blunn, Use of wind pressure coefficients to simulate natural ventilation and building energy for isolated and surrounded buildings, Build. Environ. 230 (2023) 109951. https://doi.org/10.1016/j.buildenv.2022.109951.
[21] M. Xie, J. Wang, Determination of pressure difference coefficient of shuttle elevator doors in super high-rise buildings under stack effect, Build. Environ. 232 (2023) 110076. https://doi.org/10.1016/j.buildenv.2023.110076.
[22] Z. Jiang, T. Kobayashi, T. Yamanaka, M. Sandberg, N. Choi, N. Kobayashi, K. Sano, K. Toyosawa, Wind tunnel experiment of wind-induced single-sided ventilation under generic sheltered urban area, Build. Environ. 242 (2023) 110615. https://doi.org/10.1016/j.buildenv.2023.110615.
[23] T. Van Nguyen, F. De Troyer, New surrogate model for wind pressure coefficients in a schematic urban environment with a regular pattern, Atmosphere (Basel). 9 (2018), https://doi.org/10.3390/atmos9030113.

[24] J.L.M. Hensen, on the Thermal Interaction of Building Structure and Heating and Ventilating System of Building Structure and Heating and Ventilating System, 1991.
[25] D. Cóstola, B. Blocken, J.L.M. Hensen, Overview of pressure coefficient data in building energy simulation and airflow network programs, Build. Environ. 44 (2009) 2027–2036, https://doi.org/10.1016/j.buildenv.2009.02.006.
[26] B. Knoll, J.C. Phaff, W.F. De Gids, Pressure Simulation Program, in: Updat. a Publ. 16 Th AIVC Conf., 1997: pp. 1–12.
[27] M. Grosso, Wind pressure distribution around buildings: a parametrical model, Energ. Build. 18 (1992) 101–131, https://doi.org/10.1016/0378-7788(92)90041-E.
[28] H. Montazeri, B. Blocken, CFD simulation of wind-induced pressure coefficients on buildings with and without balconies: validation and sensitivity analysis, Build. Environ. 60 (2013) 137–149, https://doi.org/10.1016/j.buildenv.2012.11.012.
[29] H. Park, D.Y. Park, Comparative analysis on predictability of natural ventilation rate based on machine learning algorithms, Build. Environ. 195 (2021) 107744. https://doi.org/10.1016/j.buildenv.2021.107744.
[30] I. Vrachimi, A.P. Melo, D. Cóstola, Prediction of wind pressure coefficients in building energy simulation using machine learning, Proc. 15th IBPSA Conf. (2017) 2467–2474.
[31] F. Bre, J.M. Gimenez, V.D. Fachinotti, Prediction of wind pressure coefficients on building surfaces using artificial neural networks, Energ. Buildings 158 (2018) 1429–1441, https://doi.org/10.1016/j.enbuild.2017.11.045.
[32] H. Wang, P. Karava, Q. Chen, Development of simple semiempirical models for calculating airflow through hopper, awning, and casement windows for single-sided natural ventilation, Energ. Buildings 96 (2015) 373–384, https://doi.org/10.1016/j.enbuild.2015.03.041.
[33] W. Wu, J.M. Han, A. Malkawi, Simplified direct forcing approach for dynamic modeling of building natural ventilation, Build. Environ. 188 (2021) 107509, https://doi.org/10.1016/j.buildenv.2020.107509.
[34] J.M. Han, Y.Q. Ang, A. Malkawi, H.W. Samuelson, Using recurrent neural networks for localized weather prediction with combined use of public airport data and on-site measurements, Build. Environ. 192 (2021) 107601.
[35] S. Charisi, M. Waszczuk, T.K. Thiis, Advances in building energy research determining building-specific wind pressure coefficients to account for the microclimate in the calculation of air infiltration in buildings, 2549 (2019). https://doi.org/10.1080/17512549.2019.1596835.
[36] A.W. Malcolm Orme, Martin Liddament, Numerical data for air infiltration and natural ventilation calculations, 1998.
[37] S. Charisi, M. Waszczuk, T.K. Thiis, Investigation of the pressure coefficient impact on the air infiltration in buildings with respect to microclimate, Energy Procedia 122 (2017) 637–642, https://doi.org/10.1016/j.egypro.2017.07.362.
[38] J.M. Han, A. Malkawi, X. Han, S. Lim, E.X. Chen, S.W. Kang, P. Howard, A two-year dataset of energy, environment, and system operations for an ultra-low energy office building, *Sci. Data* 11 (1) (2024) 938.
[39] Y. Tominaga, T. Stathopoulos, CFD simulation of near-field pollutant dispersion in the urban environment: a review of current modeling techniques, *Atmos. Environ.* 79 (2013) 716–730.