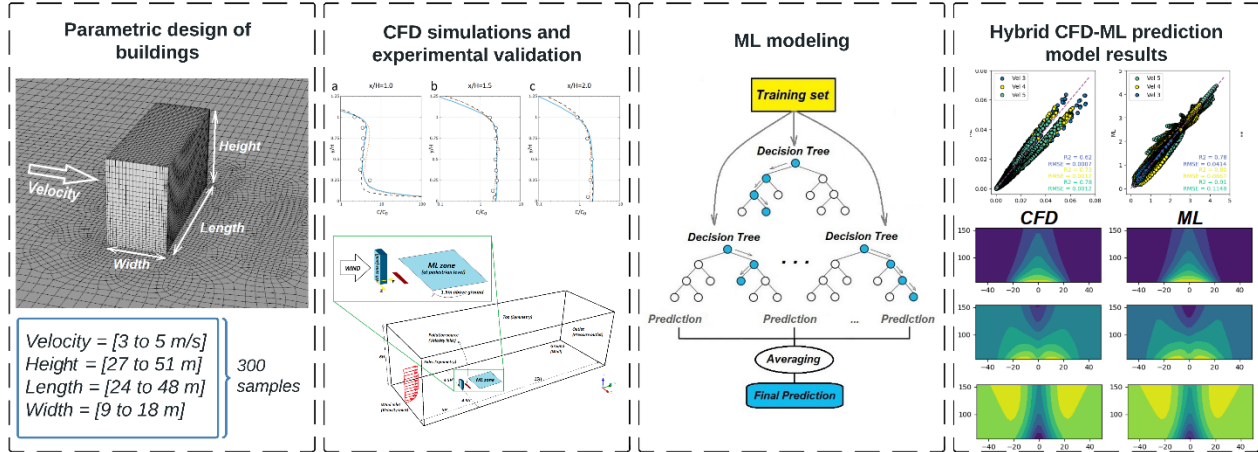


Development of Machine Learning-Aided Rapid CFD Prediction for Optimal Urban Wind Environment Design

Graphical abstract



Abstract

This paper presents a Machine Learning (ML) model based on Computational Fluid Dynamics (CFD), developed to quickly and accurately predict the impact of buildings on the urban wind environment. While CFD simulations are effective for wind studies, such as analyzing wind loads, pedestrian comfort, and pollution dispersion, they require significant computational resources and time. Recently, Machine Learning has demonstrated strong potential in providing accurate and immediate predictions by learning from datasets. By training on CFD-generated data, the ML model can quickly produce accurate and physically consistent results, addressing the limitations of CFD methods. The Reynolds-Averaged Navier-Stokes (RANS) turbulence model was chosen for CFD simulations, which were validated against experimental data, with mesh sensitivity analyzed at a wind speed of 3 m/s. A dataset of 300 cases, involving 100 hypothetical buildings and three wind speeds (3, 4, and 5 m/s), was generated to train the ML model. A multi-output regression model was proposed to effectively predict key parameters—wind velocity, turbulence intensity, and CO₂ mass fraction—in the selected urban domain. The Random Forest algorithm, which best represented the CFD results, was selected for model development. The ML model demonstrated high efficiency on new data,

achieving 88-96% accuracy. This work offers a fast and precise prediction tool, valuable for urban design and related applications.

Key words: Urban design; CFD simulation; Machine learning; Wind environment; Pollution dispersion; Pedestrian comfort

1. Introduction

Buildings and their locations significantly impact the urban wind environment and, consequently, the safety, comfort, and well-being of residents. Pre-assessments of the microclimate at building sites are essential for designing sustainable living environments. The rapid growth of cities has made urban boundary layer studies increasingly important [1]. Key factors that must be carefully investigated include wind loads on buildings, pedestrian wind comfort, and pollution dispersion. Wind loads, such as a sudden gust of wind can exert significant pressure on a building's walls, potentially causing damage. This poses a serious danger to people both inside and outside the building [2], [3]. Air pollution is a significant concern in large cities, contributing to increasing rates of respiratory diseases, allergies, and other serious health issues [4]. Adequate airflow can naturally enhance air quality, but tall buildings may obstruct it, leading to air stagnation and smog [5], [6]. When airflow collides with a building, it changes direction, creating eddies and turbulence around the structure [7]–[9]. These effects can negatively impact pedestrian wind comfort or even pose risks, such as the downdraught effect [10]–[12]. That is why it is crucial to select the optimal dimensions of a planned building in advance, thoroughly analyzing how it will influence the wind environment under various meteorological conditions.

An effective tool for this purpose is parametric analysis, which involves the use of adjustable parameters, such as dimensions, meteorological conditions, materials, and forms, to optimize architectural outcomes [13]. This method facilitates the efficient evaluation of design options tailored to specific requirements and environmental conditions, enabling in-depth analysis of energy consumption, light, and wind effects [14], [15]. Previously, such analyses were conducted using wind

tunnel tests, which involved building scale models to study the effects of varying dimensions and wind directions [16]. However, this approach was time-consuming and resource-intensive [17]. With advancements in computing technology, Computational Fluid Dynamics (CFD) has emerged as a powerful alternative, providing more precise and versatile modelling capabilities [18].

CFD is a widely used tool for modelling airflow around various geometries, enabling analysis under diverse physical conditions. It has been effectively applied to wind loads [19], urban ventilation [20], [21], pedestrian comfort, and pollution dispersion [17]. CFD modelling, based on the Navier-Stokes equations, employs turbulence models like Direct Numerical Simulation (DNS), Large Eddy Simulation (LES), and Reynolds Averaged Navier-Stokes (RANS) [22],[23]. While DNS provides high accuracy, it is computationally intensive and limited to simple geometries [24]. LES offers a balance by modelling large-scale eddies, but it remains computationally expensive [25], [26]. RANS models, with steady simulations, are approximately 80 times faster than LES [27] and are often preferred for their acceptable trade-off between accuracy and computational efficiency [28], [29]. Additionally, RANS has demonstrated strong performance in simulating flow around objects, making it the chosen approach for this study [30], [31].

Although CFD remains a powerful tool for wind environment modelling and parametric analysis [32], its computational requirements are a significant limitation, as it often takes several hours to complete each individual case simulation [33]. In contrast, an ML model trained on CFD-generated data can deliver results in just a few minutes or even seconds, while still providing highly accurate results [34]. This significant reduction in computational time demonstrates the potential of ML as a transformative tool for urban design and wind environment analysis.

ML is rapidly evolving and finding applications in a variety of fields, including speech recognition, medical diagnostics, finance, and urban design [35]. Its main advantage is its ability to process and analyze huge data sets - numbers, images, text, and audio - to identify patterns and provide optimal predictions or solutions. For urban design, ML has shown potential in optimizing energy consumption [36], reducing traffic congestion through data analytics [37], studying indoor air

quality [38], and planning energy-efficient buildings [39]. However, when applied to urban wind environments, particularly for buildings that have yet to be constructed, ML models face a significant challenge: the lack of available training data. To address this, CFD simulation provides a robust solution by generating the physically accurate data required to train ML models, bridging the gap between data scarcity and predictive capability [40].

Thus, CFD-based ML models have become increasingly relevant, as they effectively combine the precision of CFD with the computational speed of ML [41], [42]. Using data from CFD simulations, these models, once trained, deliver predictions that are both fast and physically precise, making them powerful tools for urban design and beyond [43]. In urban design, such combined models have been proposed to analyze various aspects of the urban wind environment. These applications include studying wind loads on buildings [44], [45], and solar panels [46], assessing the wind environment around buildings [47]–[49], and modelling pollution dispersion [50]–[52], including PM_{2.5} particle behaviour [53] in urban areas.

For instance, He et al. [47] developed a hybrid framework combining parametric design, CFD simulation, image processing, and ML for wind assessment around buildings, specifically identifying low-velocity areas in the wake region behind the building. Their approach relied primarily on single-output regression models and image-based feature extraction. Similarly, Du and Chen [44] applied a semi-supervised regression model for peak wind load predictions. Zhao et al. [49] introduced a two-stage CFD-Graph Neural Network (GNN) model for urban airflow prediction. Although CFD-based ML frameworks have been widely explored, most previous studies have focused on single-output models or deep-learning-based feature extraction techniques such as Convolutional Neural Networks (CNNs).

This paper proposes a comprehensive approach that integrates the assessment of wind loads, pedestrian comfort, and pollution dispersion into a single framework. Unlike previous studies that examine these aspects separately, this work considers them collectively, providing a more holistic understanding of the wind environment around buildings. This study introduces a multi-output

regression model capable of simultaneously predicting wind velocity, turbulence intensity, and CO₂ mass fraction in an urban environment. Rather than relying on image-based approximations, our approach directly extracts numerical data from CFD mesh nodes, improving both accuracy and interpretability. This also enables the prediction of full variable contours rather than only peak or minimum values. Additionally, this study integrates wind rose data specific to Almaty, Kazakhstan, providing localized meteorological inputs that enhance the relevance and applicability of the findings. Furthermore, it utilizes a synthetic dataset of hypothetical buildings generated through CFD simulations, addressing a common gap in early urban design where real-world data for unconstructed structures are often unavailable. This study also examines wind environment in the wake region behind buildings, focusing on turbulence intensity and velocity contours - an area that remains underexplored in the existing literature.

By integrating these innovative elements, this study proposes a new approach to applying CFD-based ML models in urban design, providing predictions that are not only fast and resource-efficient, but also highly accurate and adaptable to practical scenarios. This ML-assisted CFD framework can be readily applied to urban planning, helping city developers, architects, and engineers optimize building designs for improved air quality and pedestrian safety. The following sections detail the methodology, dataset preparation, and evaluation of our proposed model.

2. Development of CFD-based ML model

The development process begins with the validation of CFD simulations against experimental data, followed by a mesh sensitivity analysis (Fig.1). During this stage, the optimal RANS model and numerical settings are determined. Subsequently, parametric simulations of various building sizes are conducted using the selected RANS model at different wind velocities (3, 4, and 5 m/s), thereby

creating a training dataset for the ML model. The most appropriate ML model is then selected and trained. Finally, the ML model's accuracy is evaluated using new building sizes not included in the training data.

2.1 CFD

As previously noted, LES and DNS yield accurate results but are computationally intensive. Therefore, this paper uses a steady RANS model, known for its effectiveness in fluid flow modelling [28],[29]. Since the RANS models depend on the chosen parameterization of turbulence and the results can vary across different applications, this work considers various RANS models: standard k -epsilon model (SKE) [54], the Re-normalized Group k -epsilon model (RNG) [55] and the realizable k -epsilon model (RLZ) [56]. ANSYS Fluent software [57] was used for the simulation, where the control volume method was applied to solve the fluid flow equations. For gradient approximation, the Least Squares Cell-Based scheme was used. Advection was discretized using a second-order upwind scheme [58]. In this work, the Semi-Implicit Method for Pressure-Linked Equations (SIMPLE) was applied [59].

2.2 Benchmark test case validation

To validate the turbulence models and numerical methods, a wind tunnel experiment by Tominaga et al. [58],[60] was numerically simulated. This benchmark test effectively models airflow and pollution dispersion around a building.

(a) Computational domain configuration

The 3D computational domain of the test case is $1.08 \times 4.2 \times 1.94$ m [58], with a windward distance of $5.0H$ and a wake region length of $15.0H$ (where $H = 0.2$ m). Figure 2 illustrates the flow configuration: a cubic building model with a height H is located within the flow. A square-shaped gas source with a side length of $0.025H$ is placed at ground level behind the building. The outlet gas velocity W_S is equal to $0.5U_H$, where U_H is defined as the wind speed at height H ($U_H=0.4$ m/s). More detailed flow parameters and settings can be found in previous work [58].

(b) *Comparative analysis of numerical results with experimental data and selection of the best turbulence model*

Comparisons between the experimental data and the numerical results of various RANS k -epsilon turbulence models (RLZ, RNG, and Standard) are shown in Figs.3-4. The comparisons are performed for vertical profiles at different locations downstream of a cube at $x/H = 1.0, 1.5$, and 2.0 ($y/H = 0$), considering dimensionless velocity values U/U_H and concentration values C/C_0 . Here, C_0 represents the characteristic concentration, which is defined as follows:

$$C_0 = \frac{Q_e}{H^2 U_H} \quad (1)$$

where Q_e is the pollutant emissions flow rate, which was defined as volumetric flow rate. The value of C_0 in the experiments was 312.5 ppm (parts per million).

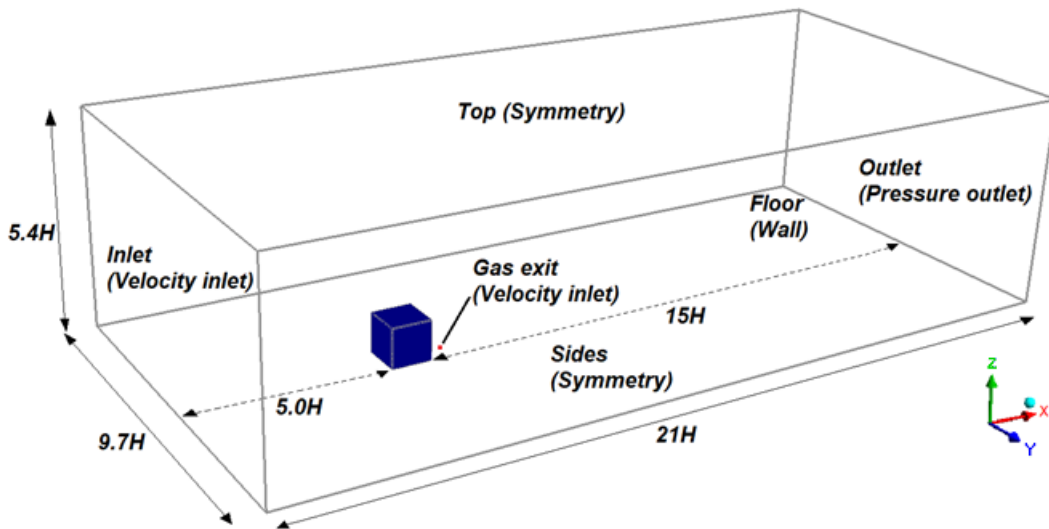


Figure 2. Computational domain and boundary conditions for the test problem [58].

There are various validation metrics for analyzing the correspondence between experimental data and simulation results: factor of 2 of the observations (FAC2), fractional bias (FB), and normalized mean square error (NMSE) [61]. These metrics are calculated as follows:

$$FAC2 = \frac{1}{N} \sum_{i=1}^N n_i \quad \text{where} \quad n_i = \begin{cases} 1 & \text{for } 0.5 \leq \frac{P_i}{Q_i} \leq 2 \\ 1 & \text{for } O_i \leq 0.05 \text{ and } P_i \leq 0.05 \\ 0 & \text{else} \end{cases} \quad (2)$$

$$FB = \frac{[O] - [P]}{0.5([O] + [P])} \quad (3)$$

$$NMSE = \frac{[(O_i - P_i)^2]}{[O][P]} \quad (4)$$

where O_i and P_i are the observed (experimental) and predicted (simulated) values of the variable for sample i , respectively; N is the number of measurement points. Square brackets indicate the average over the entire data set. Ideal values for a perfect data fit are 1.0 for FAC2 and 0 for FB and NMSE.

Table 1. Validation metrics for dimensionless concentration C/C_0 (21 sampling points)

RANS model	FAC2 (ideal value 1)	FB (ideal value 0)	NMSE (ideal value 0)
RLZ	1.0	-0.06	0.19
RNG	1.0	-0.03	0.29
SKE	1.0	-0.1	0.47

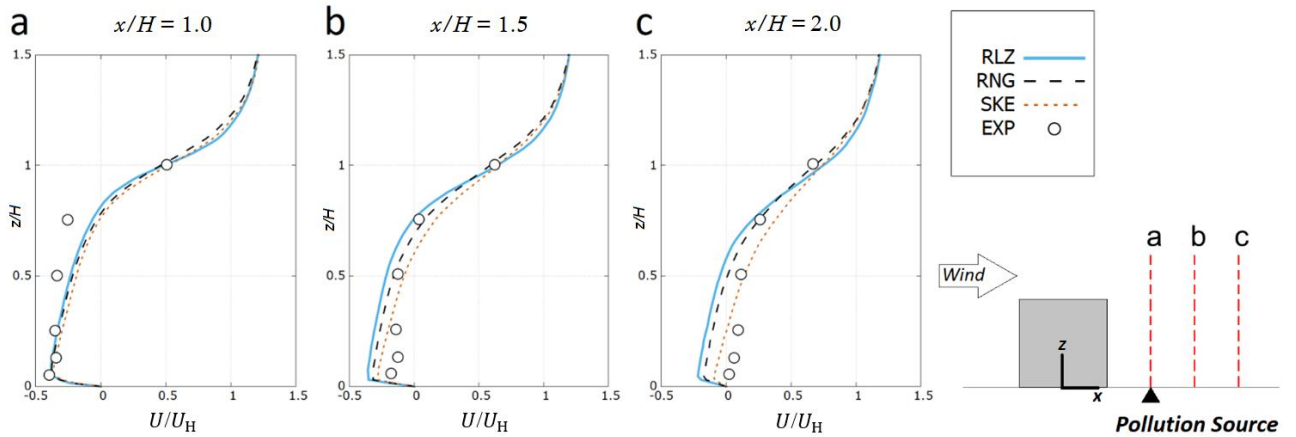


Figure 3. Comparison of the experimental data with the vertical profiles of the dimensionless flow rate U/U_H in the leeward flow region, obtained using various RANS models in the central section ($z/H = 0$): (a) $x/H = 1.0$, (b) $x/H = 1.5$, (c) $x/H = 2.0$.

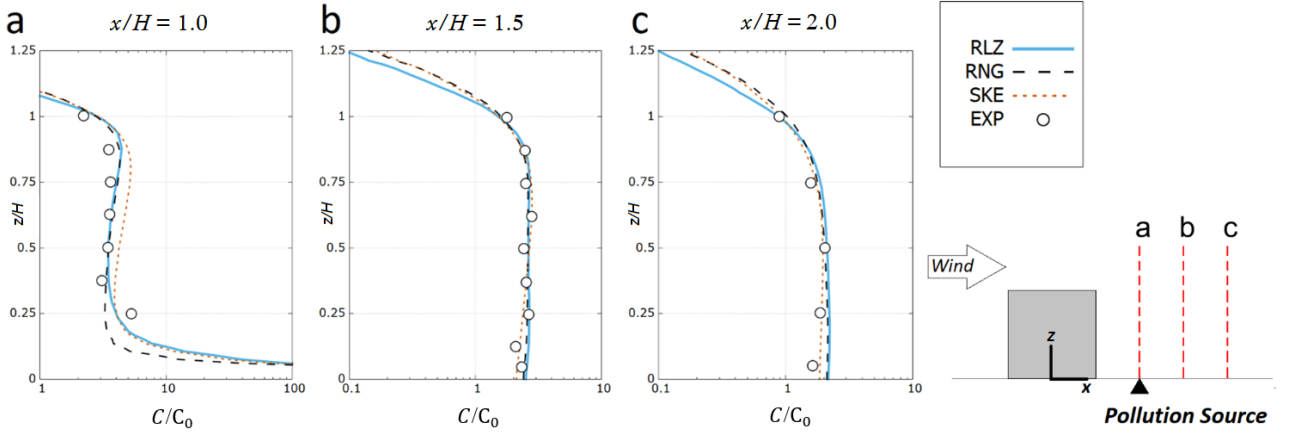


Figure 4. Comparison of the experimental data with the vertical profiles of the dimensionless concentration C/C_0 in the leeward flow region, obtained using various RANS models in the central section ($z/H = 0$): (a) $x/H = 1.0$, (b) $x/H = 1.5$; (c) $x/H = 2.0$.

Table 1 presents the FAC2, FB, and NMSE validation metrics for the vertical concentration profiles at three sections ($x = 1H, 1.5H, 2H$), based on 21 data points. All models achieved the ideal FAC2 value of 1.0, indicating good agreement with the experimental data. According to the FB metric, the RNG model performed better than the RLZ model, with a difference of 0.03 between them. However, the NMSE metric shows that the RLZ model outperformed the RNG model, with a difference of 0.1. Additionally, as shown in Fig.5, both the RLZ and RNG models show nearly the same deviation according to the FB metric across all sections, although the RNG model deviates more significantly from the experimental data than the RLZ model. Based on the comparisons of these validation metrics, the RLZ model was selected as the preferred RANS turbulence model for the simulations in this study.

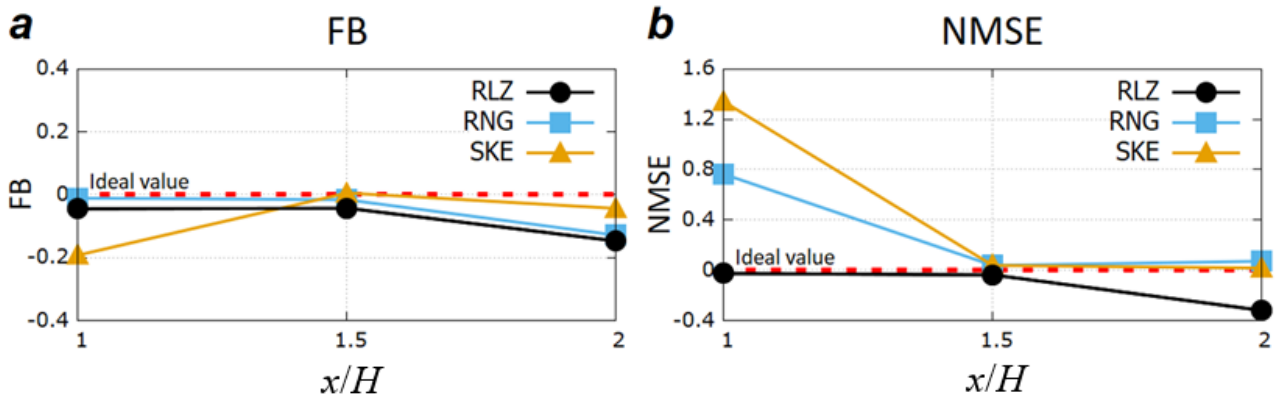


Figure 5. Comparison of RANS models at various sections ($x/H = 1.0, 1.5, 2.0$). a) FB; b) NMSE

Test studies of atmospheric flow simulations and pollution dispersion can be found in earlier investigations [62]–[68].

2.3 Main simulations: flow around building

(a) Computational domain configuration and boundary conditions

Following the successful validation of the numerical algorithm, a parametric design study was conducted. In this study the computational domain includes a building and a road as a pollution source, located 25 m behind the building (see Fig.6). The length of the road is 50 m and the width is 7 m. The dimensions of the building (height, length, and width) used as parameters and varied during the parametric analysis, which will be detailed in the next section. Following the CFD Best Practice Guideline [69], the computational domain was set up to minimize boundary effects: $5H$ from the wind inlet to the building's windward wall, $15H$ from the leeward side to the exit boundary, and $4.5H$ from the side walls. The height of the computational domain was set to $6H$. Here, H is 51 m, matching the tallest building height in the analysis.

The initial profiles of the velocity u , turbulence kinetic energy k and dissipation rate ε were established in accordance with [70], which provides suitable profiles for modelling atmospheric boundary layer, as shown in Eqs.(5a)-(5d).

$$u = \frac{u_*}{\kappa} \ln \left(\frac{y+y_0}{y_0} \right) \quad (5a)$$

$$u_* = \frac{\kappa u_{ref}}{\ln(y_{ref}+y_0)} \quad (5b)$$

$$k = \frac{u_*^2}{\sqrt{C_\mu}} \quad (5c)$$

$$\varepsilon = \frac{u_*^3}{\kappa(y+y_0)} \quad (5d)$$

where $\kappa = 0.4$, u_* is the friction velocity, y_0 – aerodynamic roughness length, set at 0.1, according to Davenport's roughness classification for open flat terrain with low crops and occasional large obstacles [71].

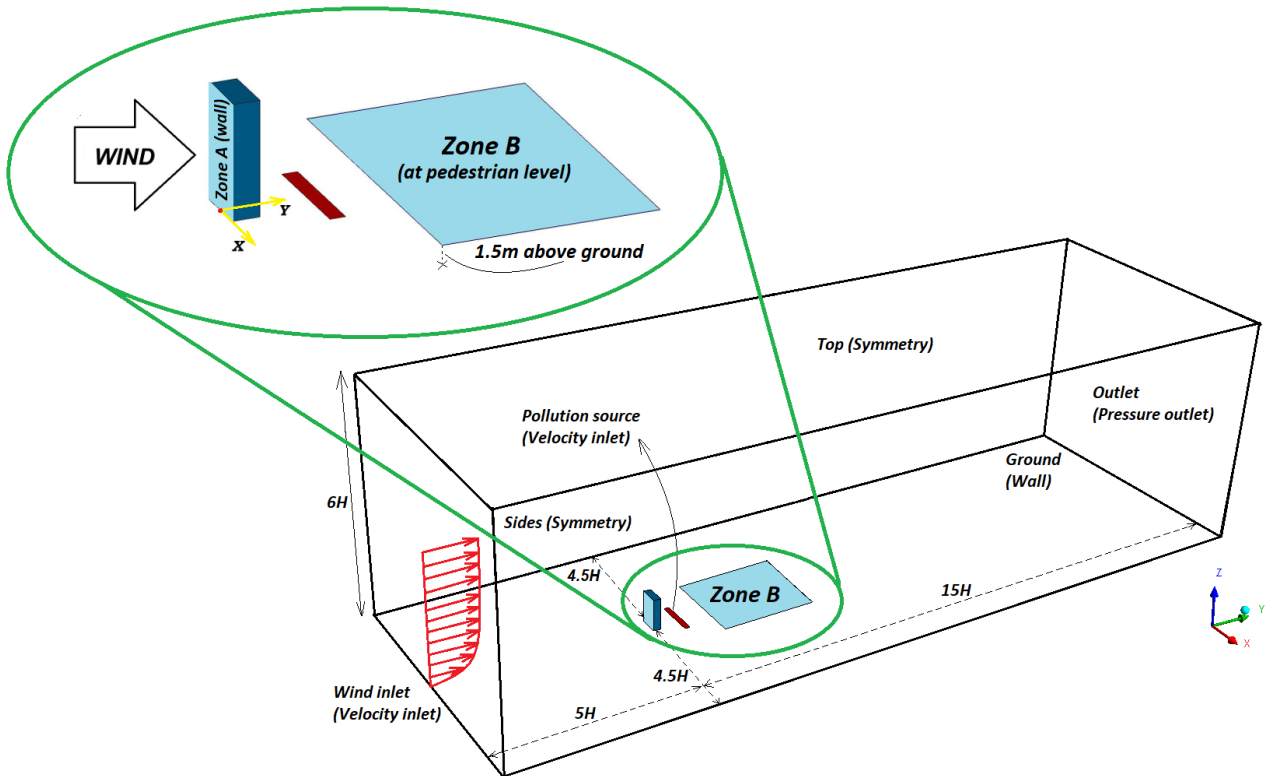


Figure 6. Computational domain, boundary conditions and ML interested zones.

This study focuses on Almaty city, Kazakhstan. According to the wind rose diagram, winds in Almaty most often blow from south to north, exceeding 1.6 m/s for 572 hours, and 3.4 m/s for 910 hours annually (Fig.7). Based on these data, the u_{ref} wind velocity parameter was set to be 3, 4, and 5 m/s, respectively [72], with $y_{\text{ref}} = 10$ m, since upper-air measurements of meteorological data are usually taken at this altitude [73]. The average road emission velocity is taken to be 1 m/s, so the momentum ratio for this model is expressed as $M = W_S/U_{10} = 0.34, 0.25, 0.2$, respectively, where W_S is the vertical exhaust velocity and U_{10} is the horizontal wind speed at a reference height 10 m. To avoid numerical errors and ensure robust convergence, the problem was solved until the flow stabilized, indicated by minimal changes in variables with each iteration. In the literature, a convergence criterion of 10^{-4} is recommended for achieving accurate solutions in urban simulation problems [74].

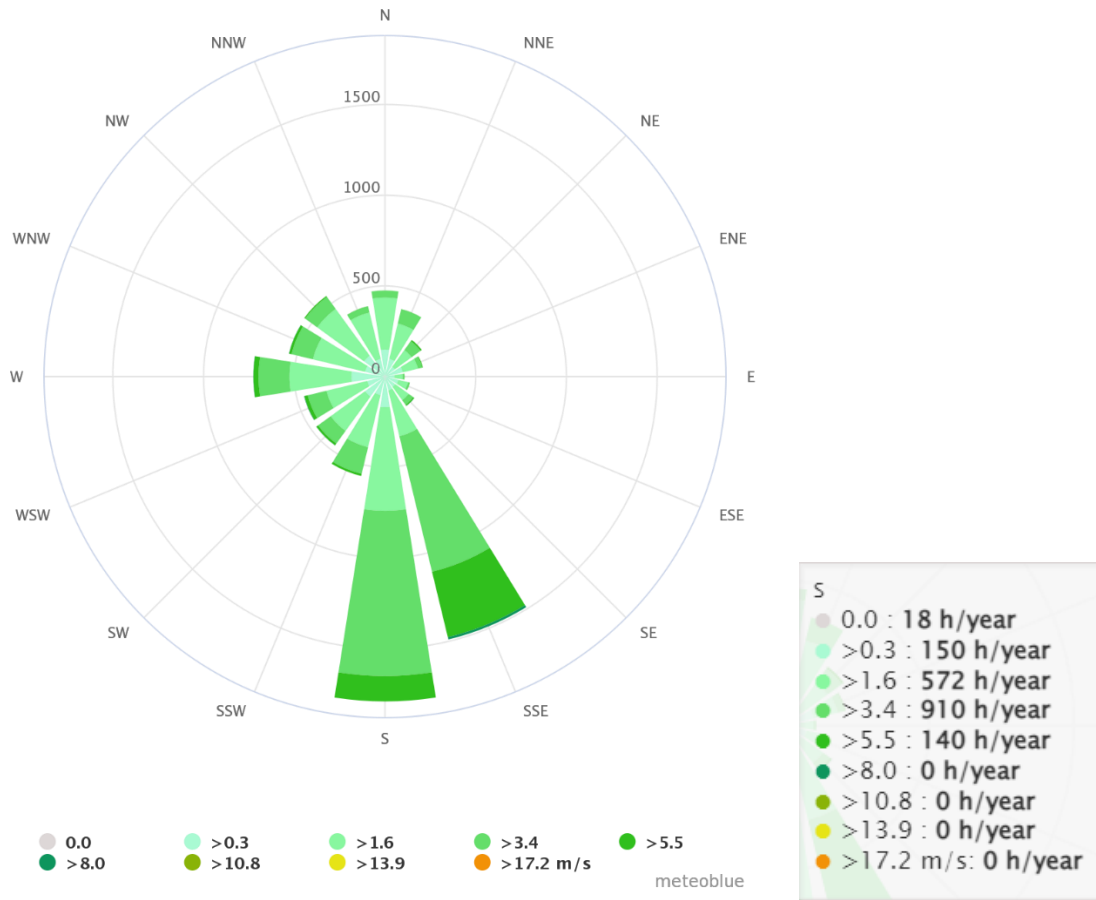


Figure 7. Wind rose diagram of Almaty city [72].

(b) *Mesh sensitivity analysis*

The computational mesh was constructed according to recommendations for detailed modelling of the atmospheric boundary layer [75]. Mesh sensitivity analysis was performed on three types of mesh: coarse (343,885 cells), basic (791,910 cells), and fine (1,165,689 cells). The number of nodes on the building walls was 2.5 million for the fine mesh, 1.0 million for the basic mesh, and 0.5 million for the coarse mesh. Mesh parameters near the building and ground were set in the following ratios: basic/coarse = 2 and fine/basic = 1.5. Simulations were performed for a building with dimensions 27 m in height, 48 m in length, and 18 m in width, with an inlet wind velocity u_{ref} of 3.0 m/s. For the mesh sensitivity analysis, velocity and turbulent kinetic energy (TKE) were considered at 2 vertical sections along the centreline: in the windward ($y = -40$ m) and in the leeward ($y = 60$ m) flow regions, with $n = 50$ points (Fig.8). The numerical results of velocity and TKE illustrate a significant difference between the basic and coarse mesh, while the results for the basic and fine mesh are almost identical.

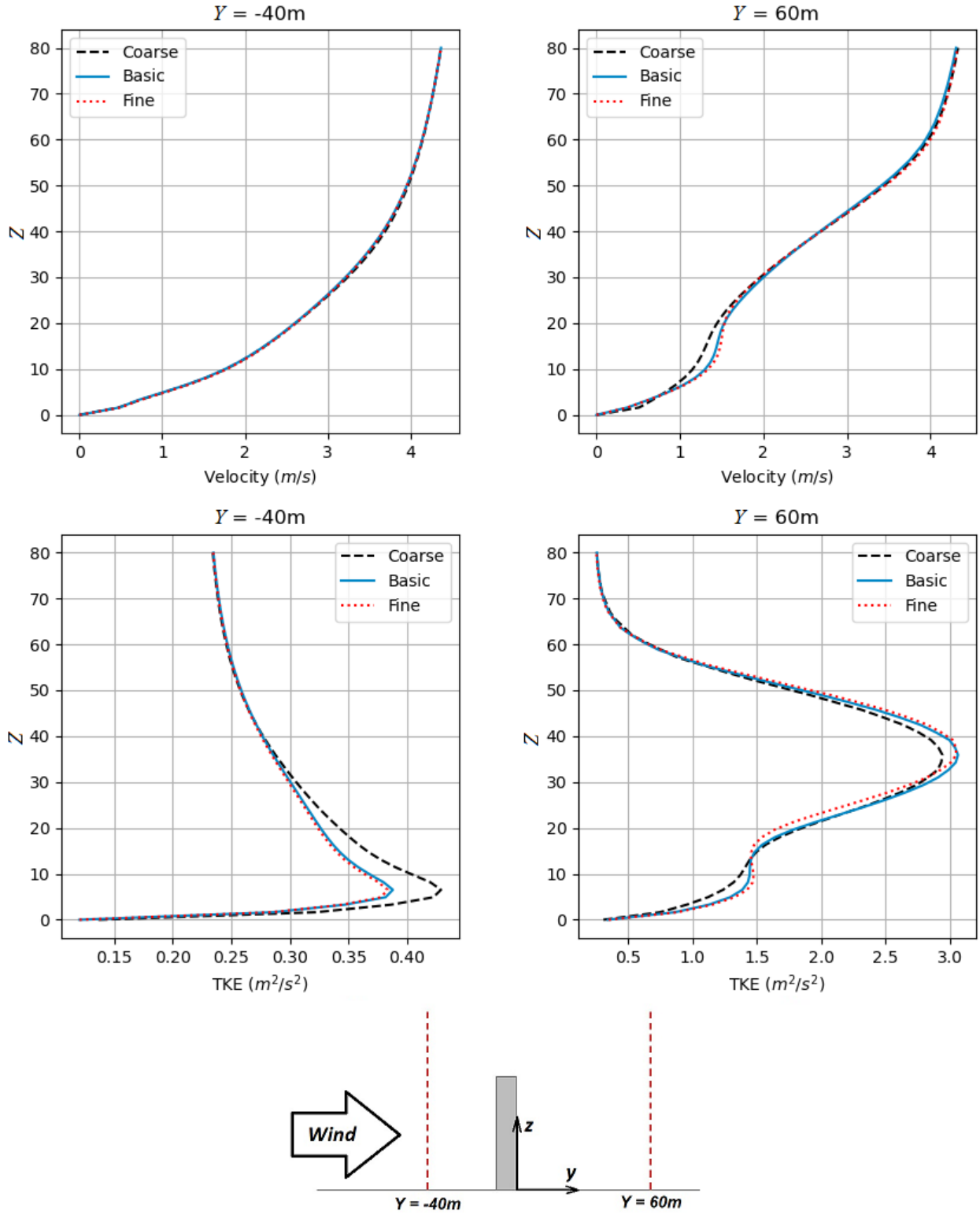


Figure 8. Comparisons of the velocity and TKE vertical profiles for different meshes in the windward ($Y = -40 m$) and leeward ($Y = 60 m$) flow regions of the building.

The mesh was evaluated using the Grid Convergence Index (GCI) presented by Richardson [76] and Roach [77]. It is calculated using Eqs.(6)-(8), where ε_{rms} is the root mean square relative error, r is the grid refinement ratio, the p parameter is based on the discretization of all second-order terms, and n is the amount of measurement points, F_S is the Safety Factor. For all three meshes, the GCI was less than 5%, aligning with the previous recommendations [78] (see table 2).

$$GCI = F_S \frac{\varepsilon_{rms}}{r^p - 1} \quad (6)$$

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

$$\varepsilon_i = \left| \frac{u_{i,coarse} - u_{i,fine}}{u_{i,fine}} \right| \quad (8)$$

where $r = 2$, $p = 2$, $n = 10$, $F_S = 1.25$.

Therefore, to optimize computational resources and reduce computation time without compromising accuracy, the basic mesh was selected for further simulations.

Table 2. Grid convergence indexes (GCI) of velocity and TKE for different meshes at two locations.

Location	Grids	Basic vs Coarse	Basic vs Fine	Coarse vs Fine
Y = 60 m	GCI of Vel (%)	1.3	0.2	1.5
	GCI of TKE (%)	2	1.5	2.9
Y = -40 m	GCI of Vel (%)	0.3	0.4	0.3
	GCI of TKE (%)	2	0.2	2.2

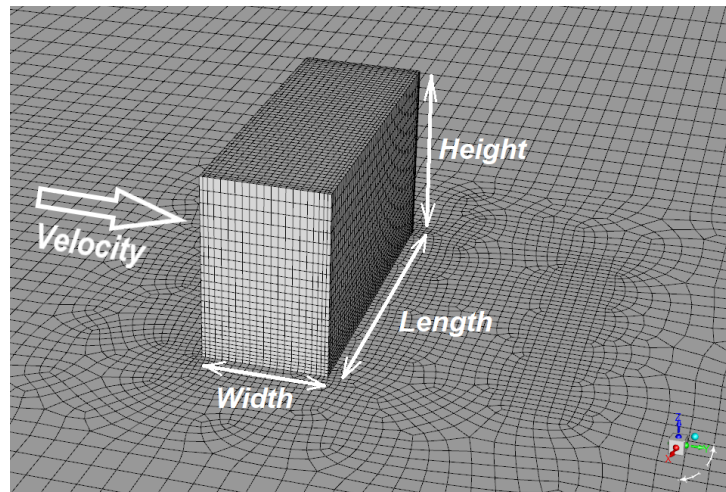


Figure 9. Building parameters and wind flow direction.

2.4. Parametric analysis

A parametric analysis was conducted involving four variables: building height, length, width, and wind velocity (Fig.9).

The choice of building height in this study was influenced by the architectural features of Almaty, where seismic safety requirements limit the construction of very high-rise buildings, resulting in a predominance of medium-height structures. Consequently, buildings ranging from 9 to 17 floors (approximately 27 to 51 meters, considering an average floor height of 3 meters) were considered. The ranges of building width and length were based on standard residential building sizes common in many modern cities. Detailed dimensions of the buildings and the selected wind speeds for parametric analysis are given in table 3.

Table 3. Building sizes and wind velocity values for parametric design.

Velocity [m/s]	Height [m]	Length [m]	Width [m]
3	27, 33, 39, 45, 51	24, 30, 36, 42, 48	9, 12, 15, 18
4	27, 33, 39, 45, 51	24, 30, 36, 42, 48	9, 12, 15, 18
5	27, 33, 39, 45, 51	24, 30, 36, 42, 48	9, 12, 15, 18

Using the selected RANS model, numerical settings, and optimal mesh configuration, 300 simulations were performed. These simulations covered 100 different building sizes, each analyzed at wind speeds of 3, 4, and 5 m/s, generating a comprehensive dataset for ML model training and validation (see table 4).

2.5. Machine Learning model development

(a) Training dataset preparation

A regression model is well-suited for the problem addressed in this work. In such tasks, the model is trained to identify relationships between input features and an output target to make predictions. These models are commonly used in tasks such as predicting apartment prices, where inputs might include area, number of rooms, and floor, with the output being the price [79]. In this study, the input

parameters were building dimensions (height, length, width) and wind velocity, while the output targets were wind velocity, turbulence intensity, and pressure coefficient. There are two types of regression models: single-output, which predicts one variable, and multi-output, which predicts multiple variables simultaneously. Both types were used in this work. The first step involved generating and processing the training data by extracting CFD simulation results from zones A and B to train the ML model (see Fig.6).

Table 4. Investigated variables and the zones from which they were extracted.

What was investigated	What variables were extracted	Datasets	From which zone	Number of outputs	ML model type
Wind load on building wall	Pressure coefficient C_p	max C_p	Zone A (wall)	1	Simple regression model (with single output)
		ave C_p		1	
		min C_p		1	
		max C_p height		1	
Pedestrian wind comfort around building	Velocity and turbulence intensity	Velocity	Zone B at pedestrian level (1.5 m height)	1156	Multioutput regression model
		Turbulence intensity		1156	
Pollution dispersion around building	CO ₂ mass fraction	CO ₂ mass fraction	Zone B at pedestrian level (1.5 m height)	1156	Multioutput regression model

Zone A was used to study the wind load on the building and represents the surface of the building's windward wall, from which the pressure coefficient (C_p) values were extracted. Zone B is a square area (99 by 99 m) located at a height of 1.5 m above the ground surface (pedestrian level) and 23 m from the road. It was used to study pedestrian wind comfort and pollution dispersion, with data on wind speed, turbulence intensity, and CO₂ mass fraction collected from this zone. The selection of this zone is justified by its practical significance, given the high pedestrian activity and the building's influence. In this region, the specified variables exhibit the most significant changes, whereas in other areas the changes are less pronounced due to their distance from the building. Moreover, when generating the mesh for all buildings, the mesh size in this particular zone remained

constant regardless of the buildings' sizes, which later facilitated the extraction of CFD results for further training of the ML model. Additionally, restricting the analysis to this specific zone significantly reduces the computational load, as considering the entire field would require processing a substantially larger volume of data.

For wind load analysis, a single-output regression model was used. Maximum, average, and minimum C_p values, along with the height of peak wind load, were extracted from Zone A, resulting in four datasets. A multi-output model was used to study pedestrian wind comfort and pollution dispersion. Training data from each grid node in the Zone B was extracted, creating 3 datasets (velocity, turbulence intensity, and CO₂ mass fraction), each with 1,156 elements. This enabled the generation of contour plots. The ML model was developed in Jupyter Lab using Python.

(b) Selection of ML models

Several tests were conducted to select the optimal ML model. The Linear Regression, Decision Tree, Random Forest, Gradient Boosting and KNN models were considered. Linear Regression is computationally efficient and interpretable, yet it assumes a linear relationship, limiting its performance with complex, non-linear data. Decision Tree models are interpretable and capture non-linear patterns, but they are prone to overfitting and exhibit high variance. Random Forest, an ensemble of decision trees, mitigates overfitting and performs well with high-dimensional data, though it is computationally intensive. Gradient Boosting offers high accuracy by iteratively combining weak learners, but it requires significant computational resources and careful tuning to avoid overfitting. K-Nearest Neighbors (KNN) is non-parametric and effective for small datasets but is computationally expensive for large datasets and sensitive to the choice of K. More details about these algorithms can be found in the ML textbook [80].

Each ML method presents unique strengths and limitations, and the choice of model depends on the specific requirements of the problem. To assess their suitability, all models were trained on the same CFD-based dataset and subsequently tested on unfamiliar input. Their accuracy was evaluated

using a range of validation metrics [81]. The R^2 metric, which indicates how well the predicted values match the known data, was used to evaluate performance:

$$R^2 = 1 - \frac{\sum(y_{true} - y_{pred})^2}{\sum(y_{true} - y_{mean})^2} \quad (9)$$

where y_{true} are the known values, y_{pred} are the values predicted using ML, y_{mean} is the mean value of the known values. The value of R^2 ranges from 0 to 1, with 1 indicating perfect agreement between predicted and actual values. Based on Table 5, the Random Forest model outperformed the others in terms of predictive accuracy with an R^2 value of 0.885, RMSE of 0.001, and MAPE of 7.68%, which are the closest to the ideal values. Additionally, it offered a reasonable balance between training time (30.4 seconds) and prediction time (96 milliseconds), making it both efficient and accurate. For these reasons, the Random Forest model was selected for further simulations.

Table 5. Comparison of different ML algorithms: performance on unknown dataset

ML algorithm	R^2 (ideal value: 1.0)	RMSE (ideal value: 0.0)	MAPE (ideal value: 0%)	Training time/ prediction time
Linear Regression	0.654	0.0031	25.21%	7.36 sec/ 91.6 ms
Decision Tree	0.882	0.0011	7.72%	7.15 sec/ 101 ms
Random Forest	0.885	0.001	7.68%	30.4 sec/ 96 ms
Gradient Boosting	0.867	0.0011	11.45%	143.95 sec/ 108.3 ms
KNN	0.614	0.0035	42.14%	9.93 sec/ 147 ms

The Random Forest algorithm operates by generating multiple data subsets from the original training set through bootstrapping [82]. For each subset, a distinct decision tree is constructed and trained on its unique sample of data. Each decision tree independently generates a prediction. Once all trees have made their outputs, these are combined through an averaging process to yield the final prediction (Fig.10). This aggregation from multiple trees enhances the stability and accuracy of the model.

Several key parameters are used in this model:

- Number of trees ($n_estimators$) – Provides a balance between computational cost and prediction accuracy, as increasing the number of trees can improve stability but also increases training time.
- Maximum depth (max_depth) – Controls the complexity of each tree, preventing overly deep structures that could lead to overfitting.
- Minimum samples per leaf ($min_samples_leaf$) – Allows each leaf node to represent individual samples, capturing detailed patterns in the data while keeping the model simple.
- Minimum samples per split ($min_samples_split$) – Ensures that each decision node is sufficiently representative.

The following key structural parameters were used in this study: $n_estimators = 10$, $max_depth = 10$, $min_samples_leaf = 1$, and $min_samples_split = 2$. These parameters were selected to achieve a balance between computational efficiency and model accuracy.

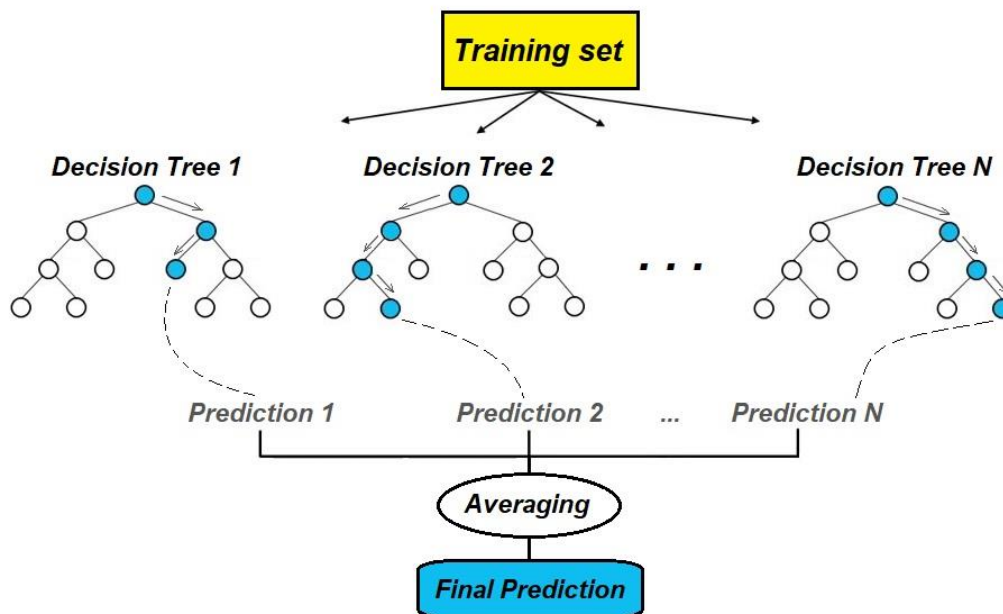


Figure 10. Schematic of the Random Forest ML algorithm

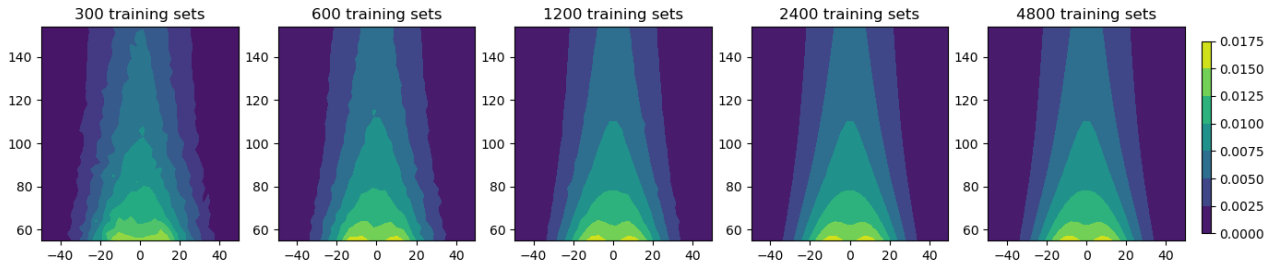


Figure 11. CO₂ mass fraction predicted by ML models based on different training dataset size.

(c) Structural analysis

The next phase aimed to improve the ML model's performance by evaluating the sufficiency of the training data size. Figure 11 shows that the initial dataset of 300 cases was insufficient, as indicated by the "rough" plot. To address this, the dataset size was increased through duplication. Table 6 presents the dataset sizes, training accuracy, and Mean Absolute Percent Error (MAPE) on the validation set for CO₂ dispersion. The results suggest that a dataset size of 4,800 is optimal for training the Random Forest model. The effectiveness of data duplication for improving model performance is particularly notable in Random Forest models. The more frequently these input-output relationships occur, the better the ML can learn the patterns. However, it is important to note that this approach is specifically useful for Random Forest models and may not be suitable for other ML algorithms. In other models, data duplication can lead to overfitting, improving performance on the training set but potentially degrading accuracy on unfamiliar datasets.

Table 6. Performance of ML models based on different training datasets: R^2 training scores and Mean Absolute Percentage Error (MAPE) compared with CFD simulation results.

Sizes of training dataset	R^2	MAPE
300	0.9931075	0.1914
600	0.9986719	0.1679
1200	0.9998891	0.1646
2400	0.9999982	0.1643
4800	1.0	0.1643

3. Results

The trained ML model was tested on a new dataset of eight buildings with unknown dimensions different from the training set but within its size range (see table 7).

Table 7. Building sizes for the new unknown test dataset

No. of test buildings	Height [m]	Length [m]	Width [m]
1	37	20	9
2	29	21	10
3	47	25	16
4	31	27	16
5	43	30	11
6	35	44	10
7	41	45	11
8	49	47	17

Each building was evaluated under three wind speeds (3, 4, and 5 m/s), totalling 24 test cases. The ML model's predictions were compared with CFD simulations, showing a high correlation (R^2 ranging from 88% to 96%) and confirming the model's accuracy in predicting the wind environment around buildings. For a more detailed analysis, table 8 shows the values of other statistical metrics (FAC2, FB, NMSE and R^2 , as calculated by Eqs.(2-4) and (9)).

The CFD simulations for each building required an average of 3 - 4 hours to complete, totalling approximately 1.5 months to generate the full set of 300 CFD training data for the ML model. In contrast, the training phase of the ML model was completed in just 30.4 seconds, with a prediction time of 96 milliseconds (ms) per case. The ANSYS simulations were conducted on a high-performance desktop featuring an 8-core Intel Core i7-9700KF processor and 32 GB of RAM, while the Machine Learning tasks were performed on a laptop equipped with an Intel Core i7-1065G7 processor and 16 GB of RAM. Despite running on a less powerful machine, the ML model demonstrated remarkable efficiency and achieved significant time savings.

Detailed results are discussed below.

Table 8. Comparison metrics between ML and CFD models

Variables	FAC2 (ideal value 1)	FB (ideal value 0)	NMSE (ideal value 0)	R^2 (ideal value 1)
Velocity	1.0	0.0058	0.002	0.96
Turbulence intensity	1.0	0.0004	0.0003	0.99
CO ₂	1.0	0.0617	0.1167	0.88

3.1 Wind Load on building (Zone A)

Figures 12 and 13 compare wind load results from ML and CFD. Figure 12 highlights the maximum C_p and the height of peak wind pressure on the building wall. The height of the peak pressure point did not depend on wind strength but only on the building's height: the taller the building, the higher the point of maximum pressure. This indicates that the upper floors of buildings experience more wind pressure than the lower floors. It is evident from the plots that wind speed significantly influences the pressure force. Specifically, at 3 m/s, the maximum C_p does not exceed 17; at 4 m/s, it ranges from 22 to 30; and at 5 m/s, it exceeds 44. The R^2 value ranged from a minimum of 86% to a maximum of 93%. Figure 13 shows the minimum and average pressure coefficient C_p .

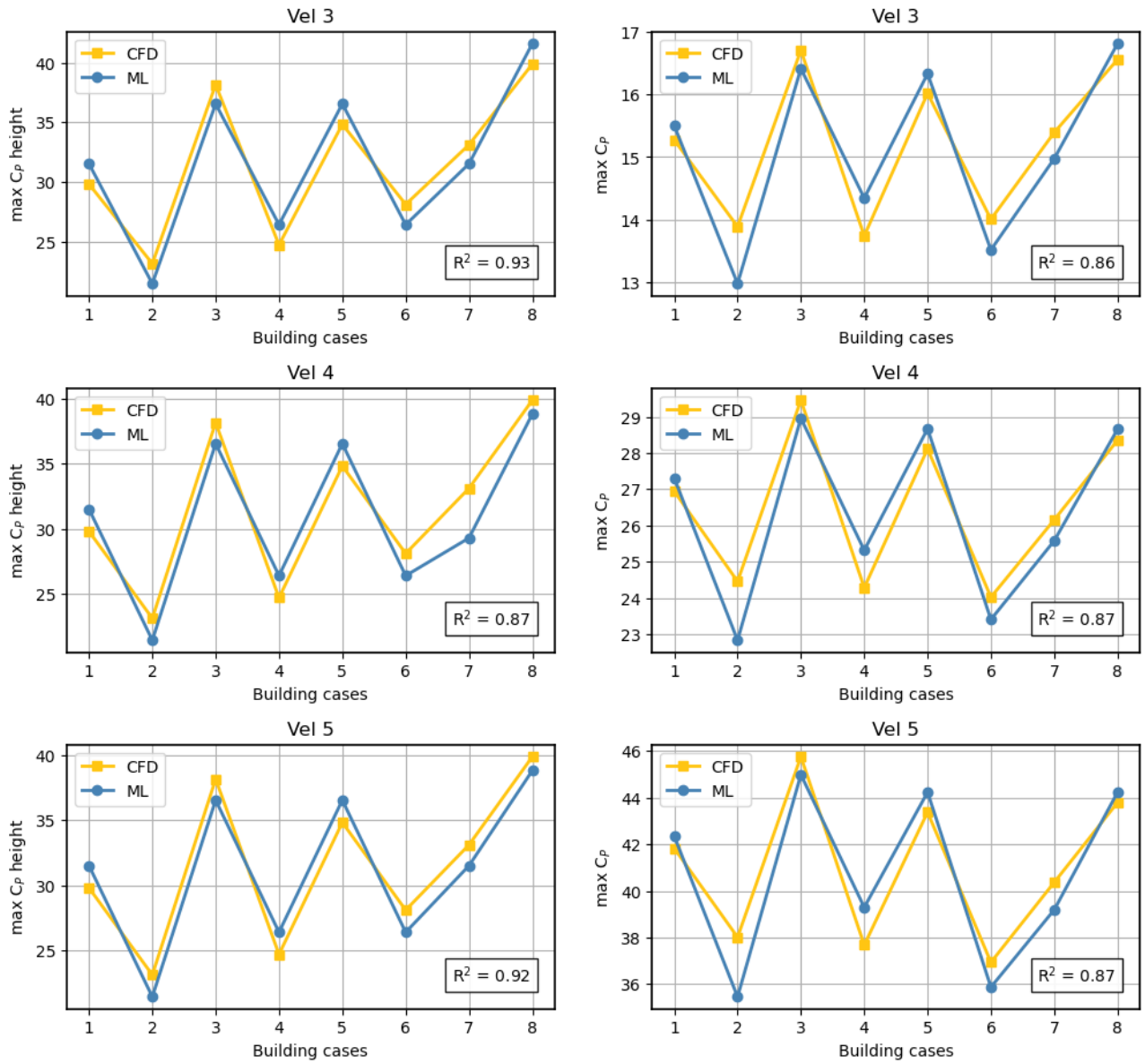


Figure 12. Comparison of CFD and ML models for the maximum C_p [-] (peak wind loads) and their height position [m].

3.2 Pedestrian wind comfort and pollution dispersion (Zone B)

The ML model's accuracy in predicting CO_2 mass fraction, velocity, and turbulence intensity at pedestrian level (Zone B) was evaluated against CFD results. Figure 14 presents scatter plots of these variables at 1,156 points (the number of nodes in Zone B). Overall, there is a high degree of similarity between the methods, though some deviations are observed at individual points. Depending on wind speed, R^2 metrics ranged from 62% to 78% for CO_2 mass fraction, 78% to 91% for velocity and 87% to 92% for turbulence intensity. Root Mean Square Error (RMSE) values were also calculated,

showing close agreement between the models: for CO₂ mass fraction it ranged from 0.0007 to 0.0012; for velocity from 0.0414 to 0.1148; and for turbulence intensity from 0.0063 to 0.012.

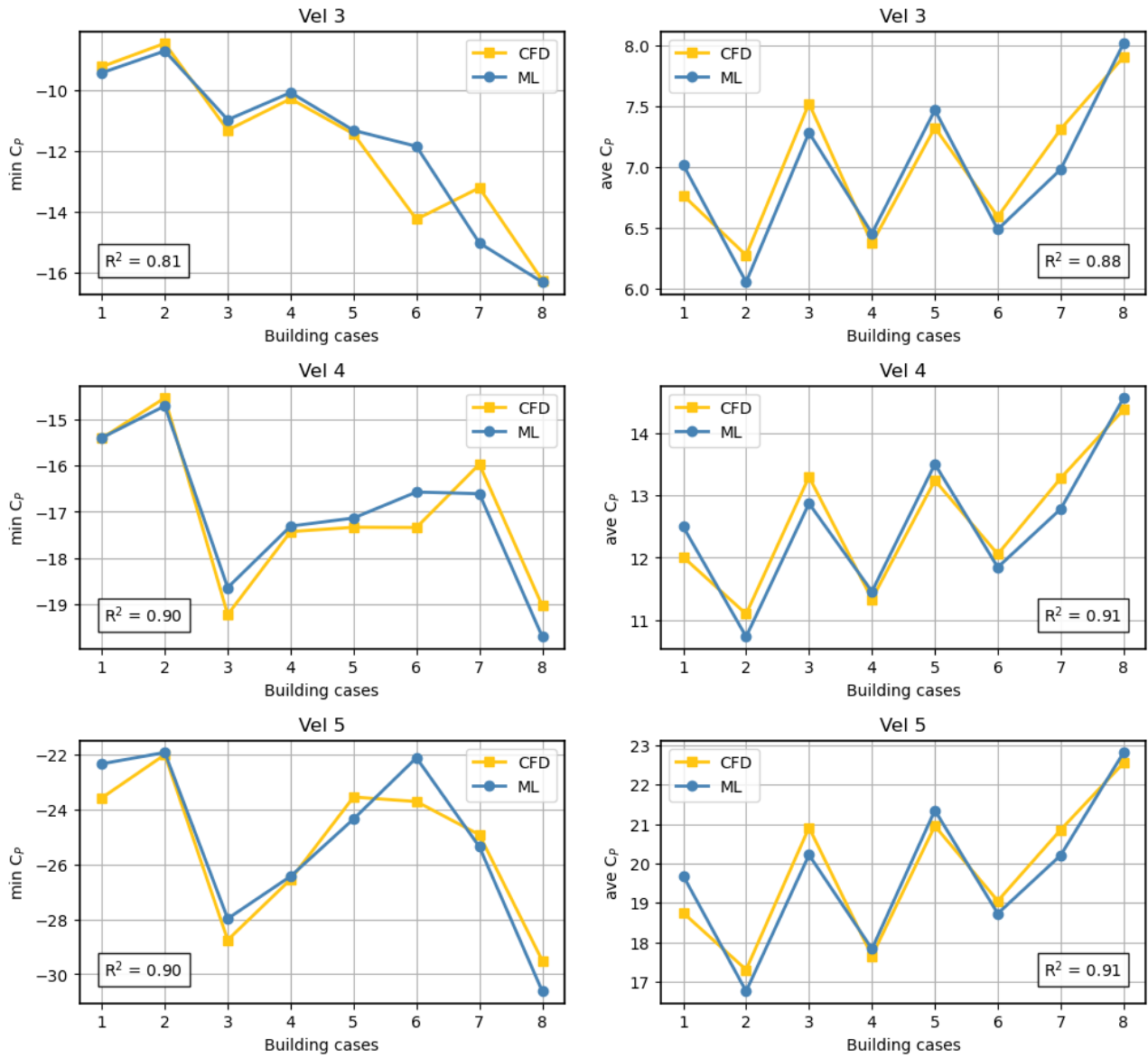


Figure 13. Comparison of CFD and ML models for minimum and average C_p .

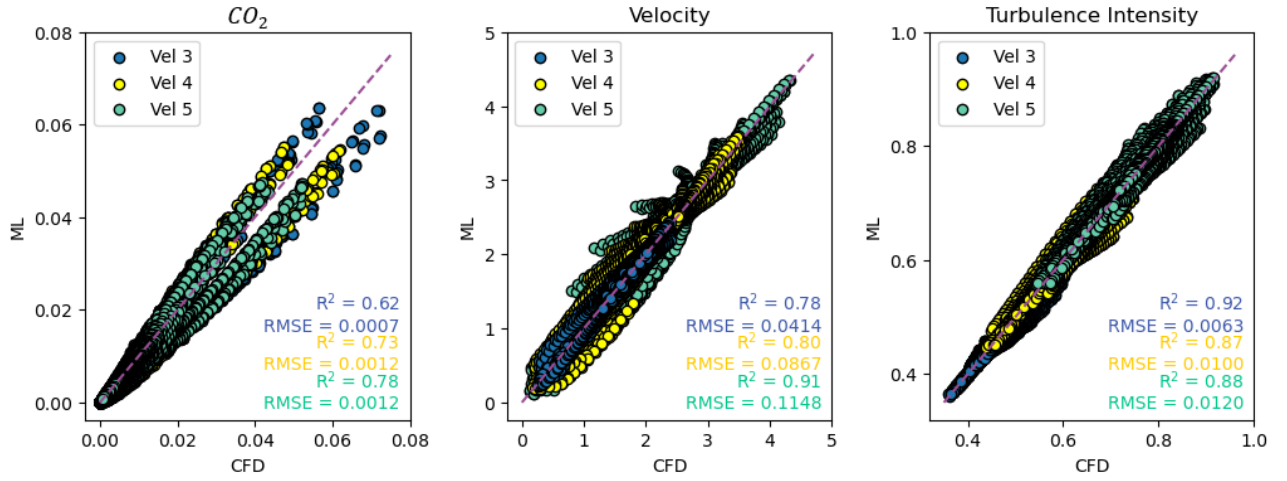


Figure 14. Comparison of CO_2 mass fraction, velocity, and turbulence intensity between CFD and ML results: scatter plots and metrics.

Figures 15-23 illustrate 2D contours taken from Zone B. These results also show excellent agreement between ML and CFD predictions. In Fig.15, for a wind speed of 3 m/s, the maximum flow velocity at pedestrian level is 2.7 m/s for test building number 8 (see table 7). In other cases, it remains below 2.5 m/s. At a wind speed of 4 m/s, building 8 again experiences the highest air flow at pedestrian level, reaching 4 m/s, while in other buildings it stays below 3.2 m/s (Fig.16). With a wind of 5 m/s, the maximum flow velocity reaches 4.8 m/s for building 8, 4 m/s for buildings 5-7 and does not exceed 3.6 m/s for others (Fig.17).

The influence of building length on the flow pattern behind the building for wind speeds greater than 4 m/s is demonstrated in Figs.16-17. The longer the lee wall of the building (increases from the 1st to the 8th test building), the larger the low-velocity areas in leeward zone. It should be noted that in cases 5 - 8, where the wall length is significant, the speed behind the building is very low (Fig.17), but at the same time the speed at the edges of the building is higher due to wind flowing around the building, creating eddies and turbulence. According to the Lawson Wind Comfort Criteria, pedestrian-level air velocity should not exceed 7.6 m/s [83]. Although the flow in these cases stays below this critical value, it is crucial to consider it for scenarios where nearby buildings might create high-velocity zones (Bernoulli effect). Turbulence intensity increases with wind speed, as expected:

at 3 m/s, the maximum value does not exceed 0.8, at 4 m/s, it reaches 0.85, and at 5 m/s, it goes up to 0.95 (Figs.18-20).

Figures 21–23 show that higher wind speeds improve pollution dispersion, with the maximum CO₂ mass fractions decreasing from 0.064 at 3 m/s to 0.056 at 4 m/s and 0.048 at 5 m/s. Therefore, maintaining adequate ventilation in urban streets and preventing the formation of low-velocity regions are crucial for ensuring effective pollutant dispersion and improving air quality.

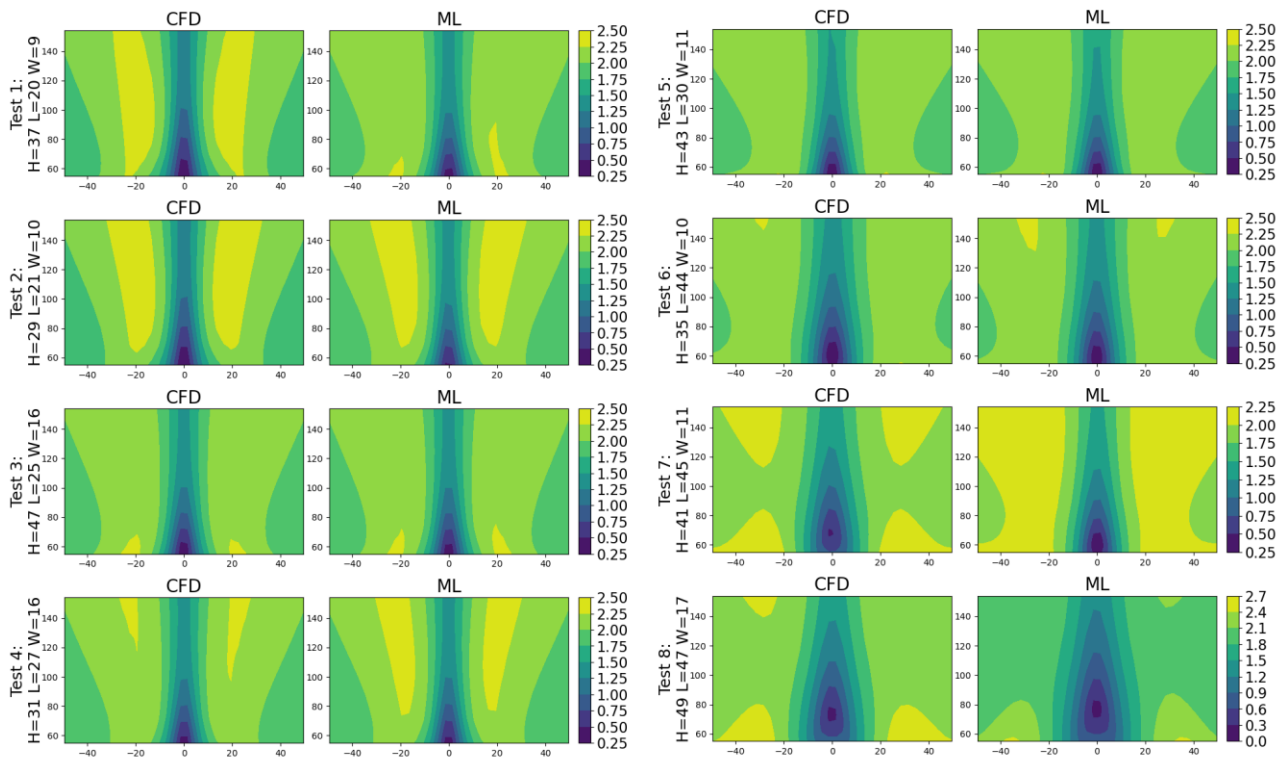


Figure 15. Velocity contour plots: comparison between CFD and ML results (wind velocity 3 m/s).

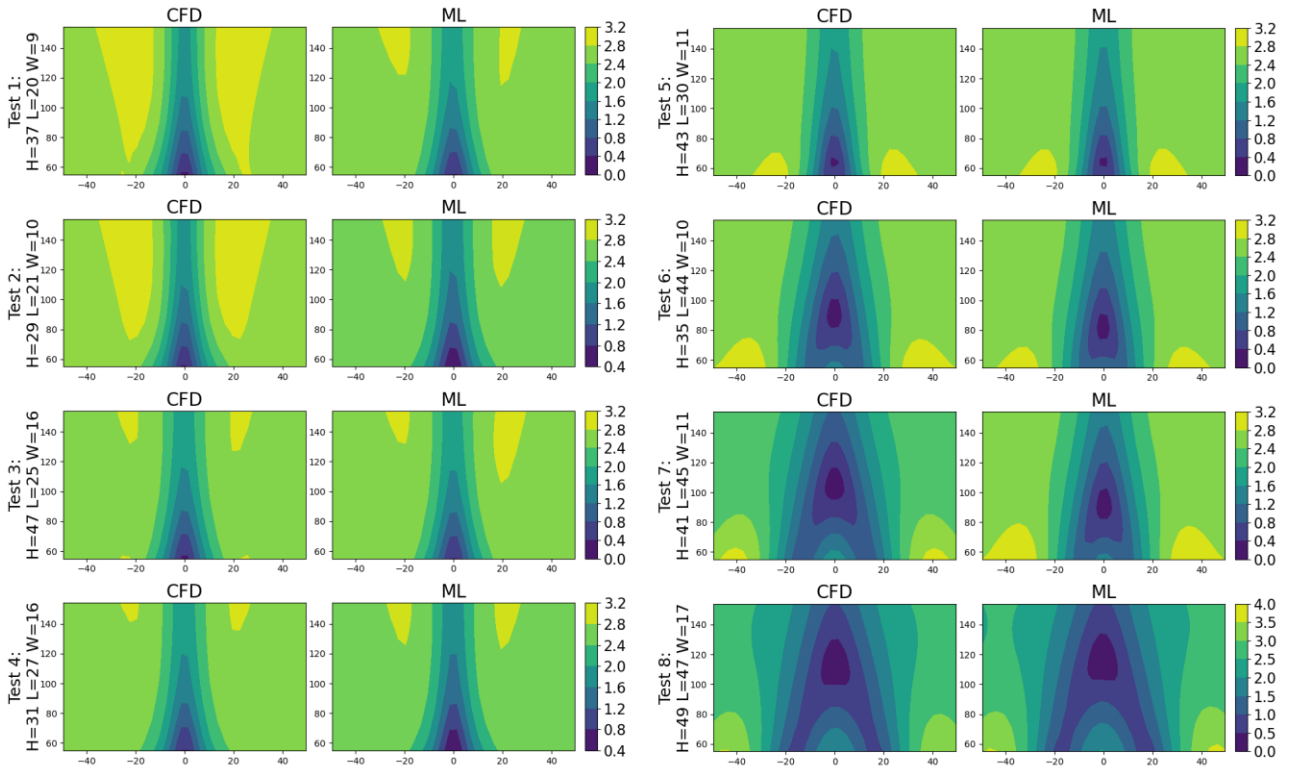


Figure 16. Velocity contour plots: comparison between CFD and ML results (wind velocity 4 m/s).

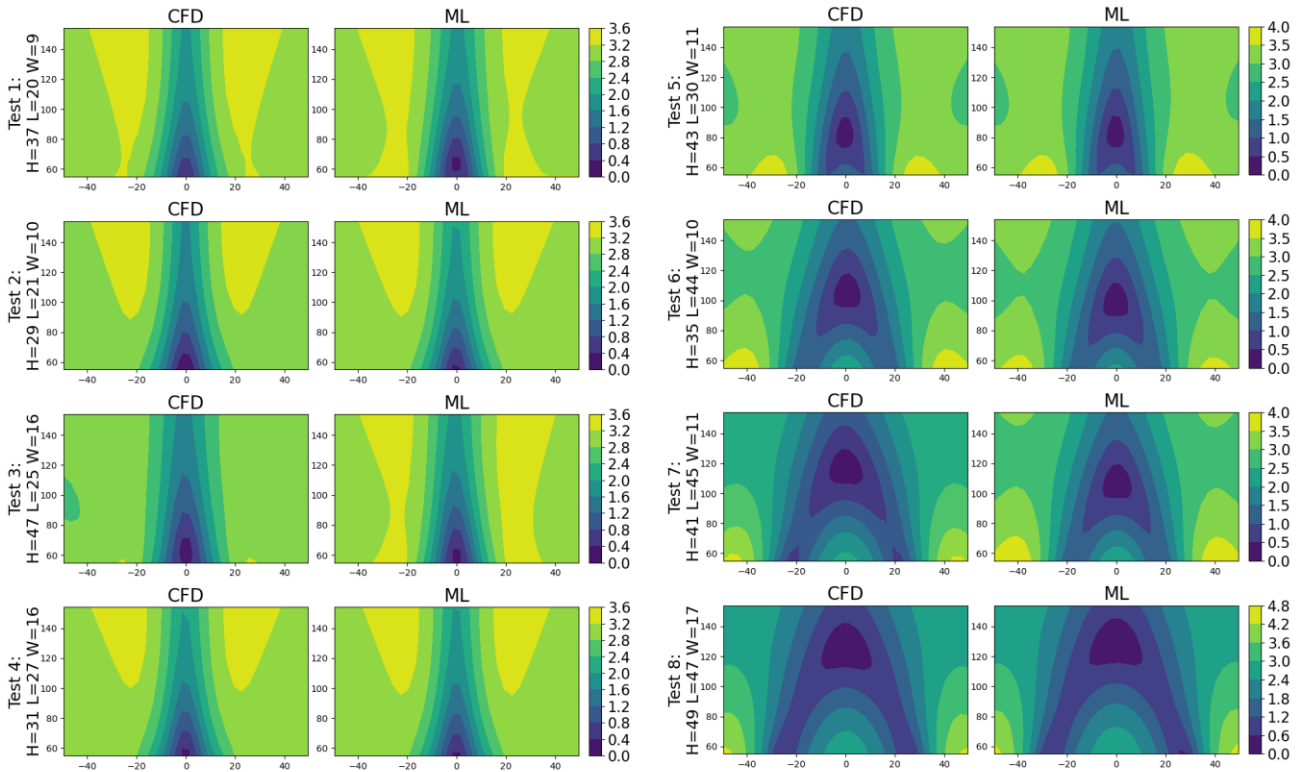


Figure 17. Velocity contour plots: comparison between CFD and ML results (wind velocity 5 m/s).

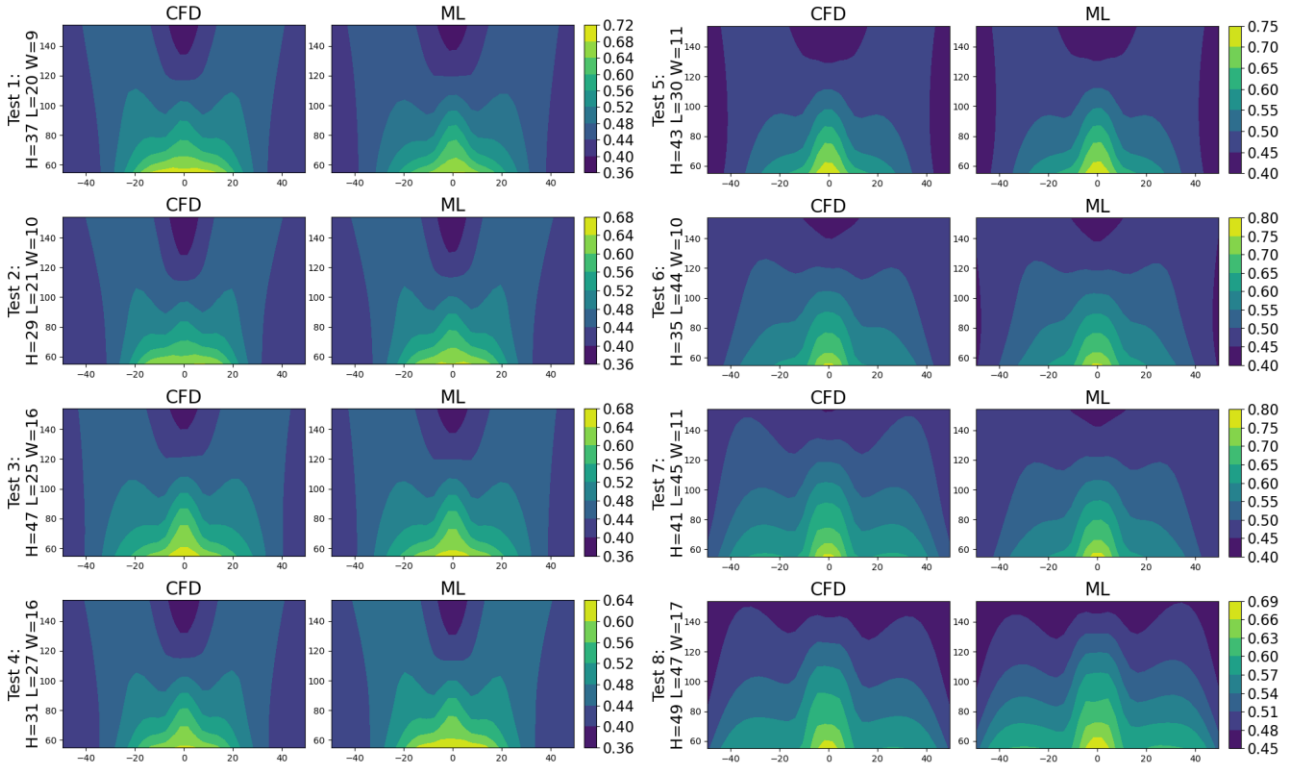


Figure 18. Turbulence intensity contour plots: comparison between CFD and ML results (wind velocity 3 m/s).

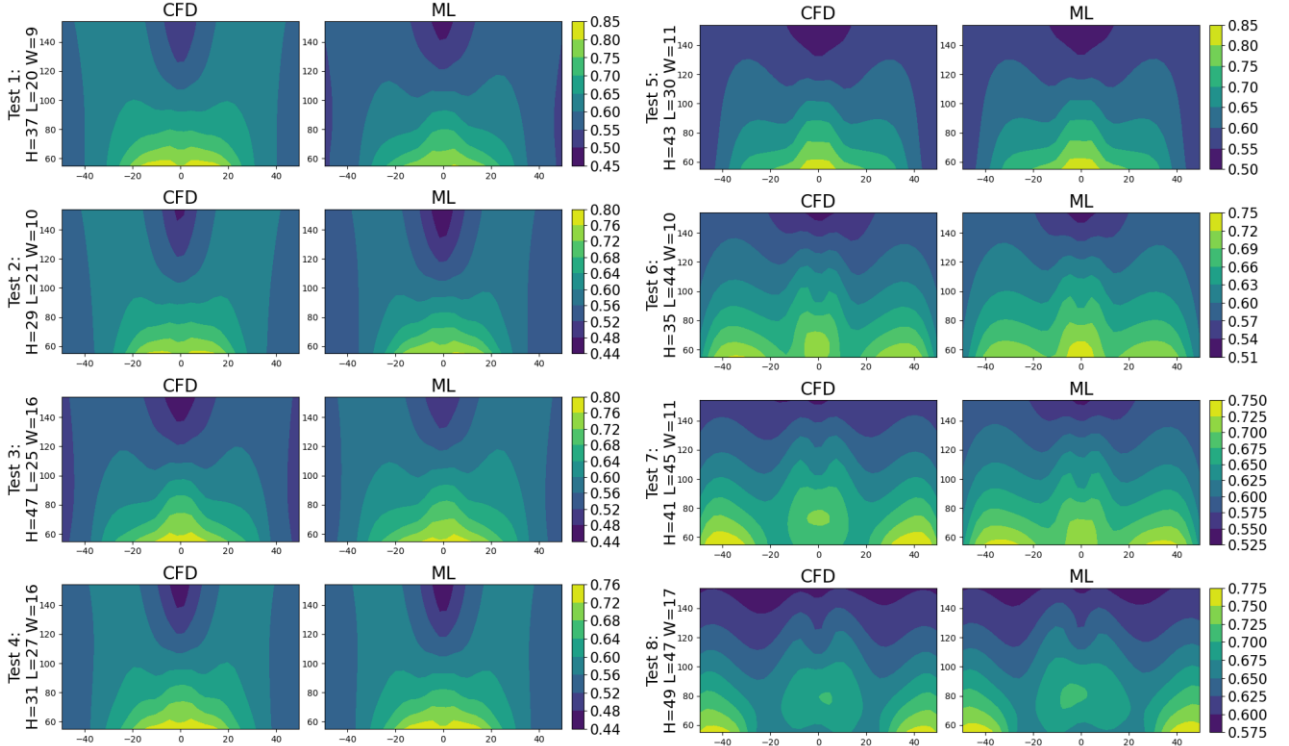


Figure 19. Turbulence intensity contour plots: comparison between CFD and ML results (wind velocity 4 m/s).

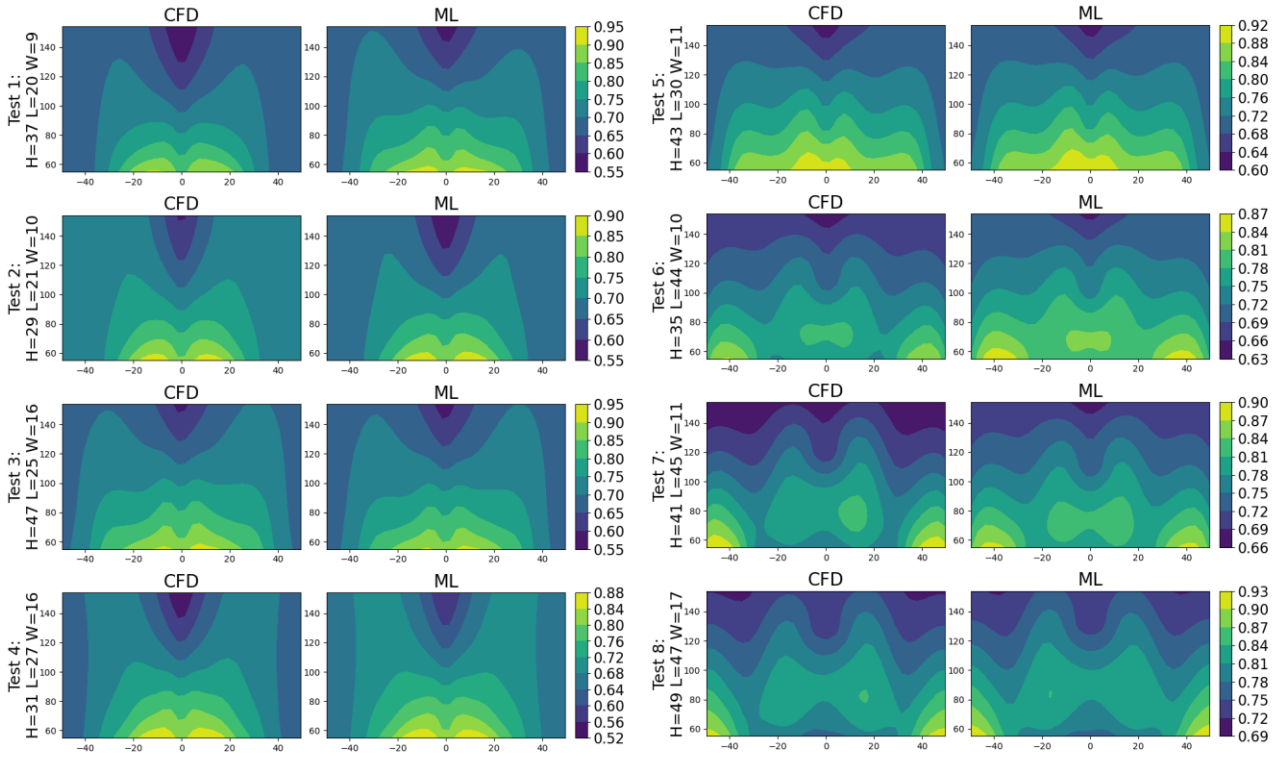


Figure 20. Turbulence intensity contour plots: comparison between CFD and ML results (wind velocity 5 m/s).

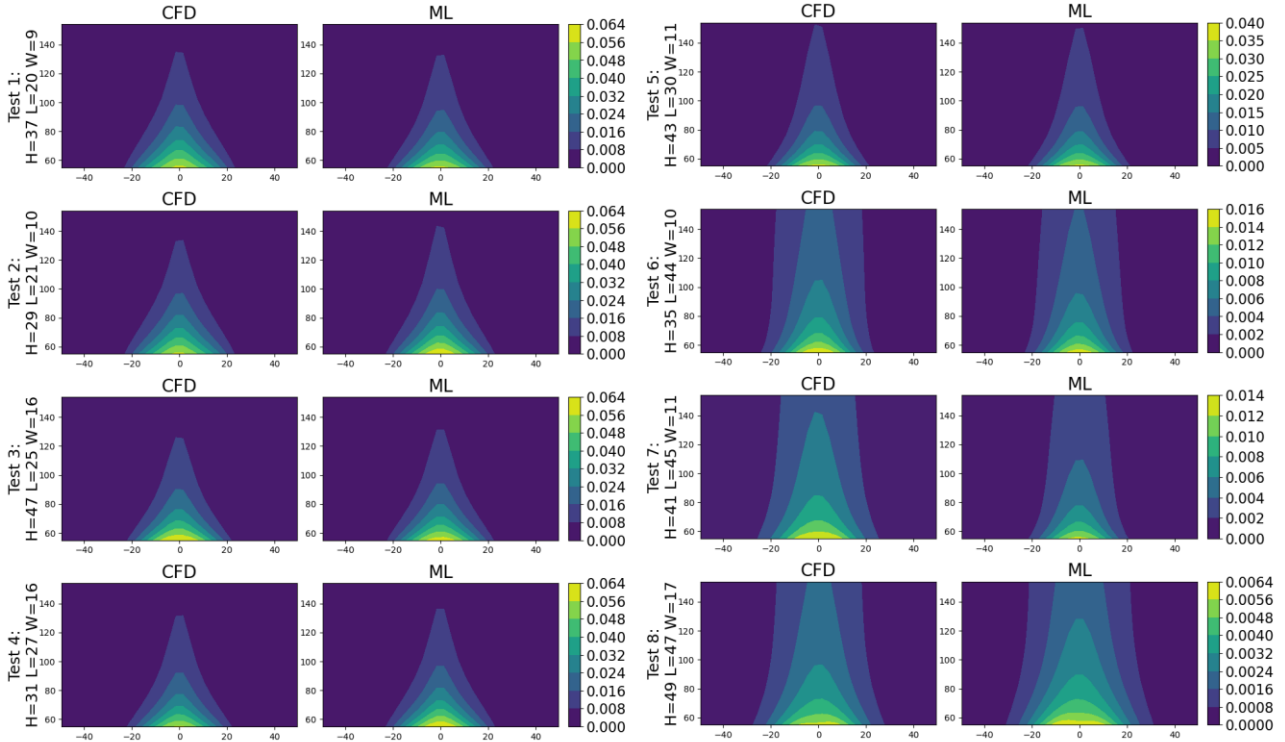


Figure 21. CO₂ mass fraction contour plots: comparison between CFD and ML results (wind velocity 3 m/s).

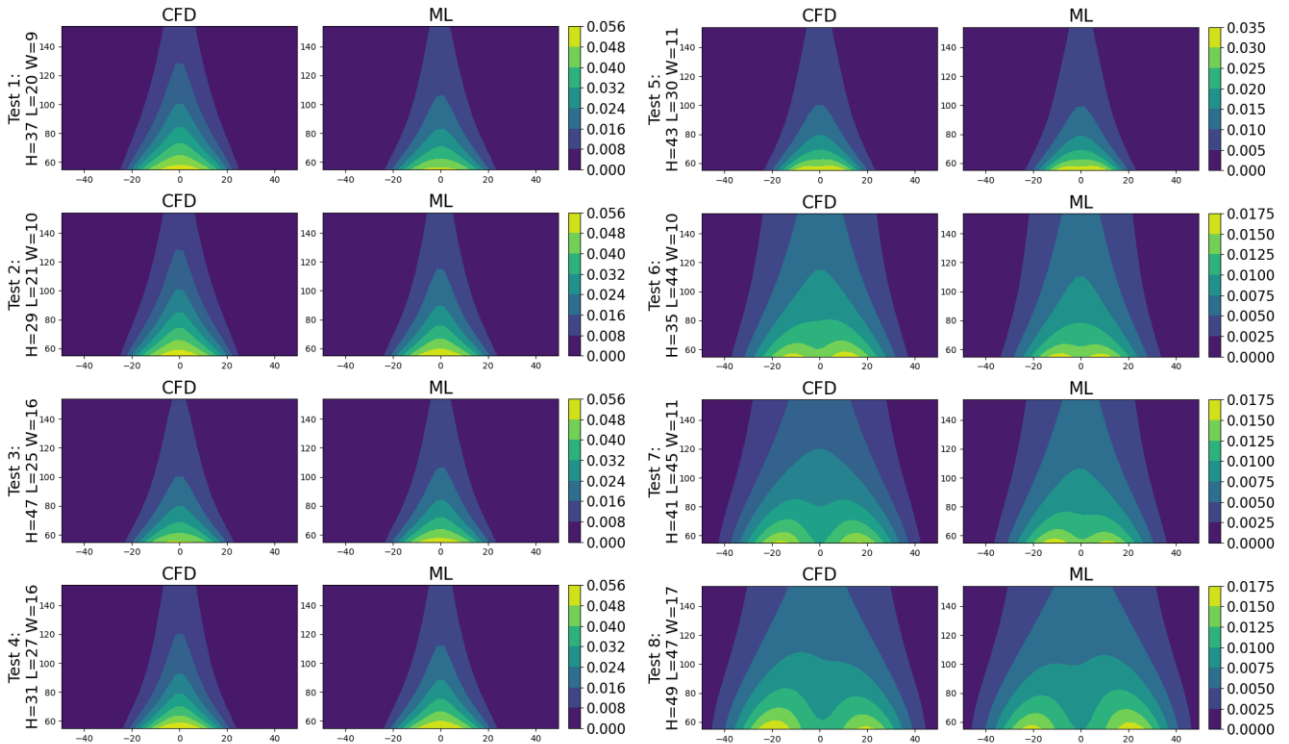


Figure 22. CO₂ mass fraction contour plots: comparison between CFD and ML results (wind velocity 4 m/s).

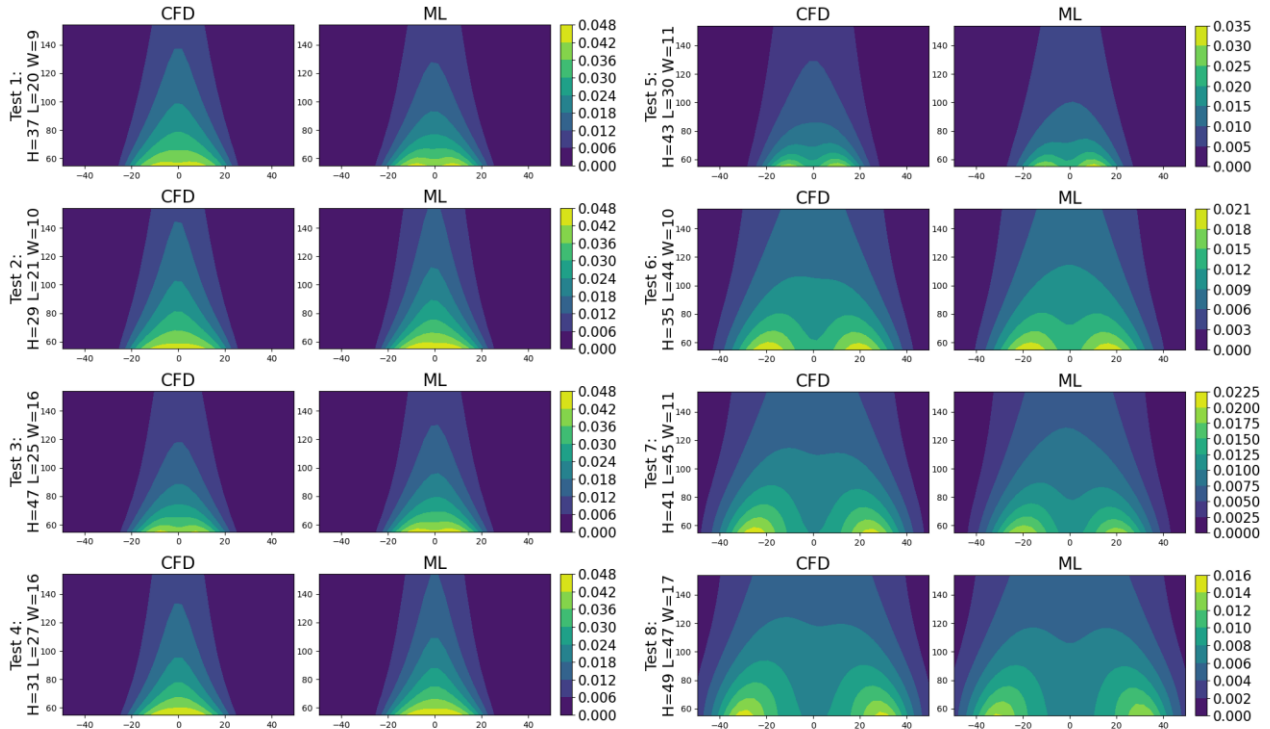


Figure 23. CO₂ mass fraction contour plots: comparison between CFD and ML results (wind velocity 5 m/s).

4. Discussion

Model advantages

The CFD-based ML model proposed in this study offers several advantages. It simplifies the process by bypassing complex steps in CFD simulations, such as geometry construction, boundary condition definition, mesh generation, and post-processing. Instead, inputting the building dimensions and wind velocity directly yields accurate predictions, significantly streamlining the workflow. This model also provides substantial time efficiency, delivering results within seconds compared to the days or weeks required by traditional CFD methods, allowing for rapid identification of optimal designs. The proposed method also reduces the need for extensive simulations, leading to significant cost savings in electricity, equipment, and experimental setups. Pre-simulation of design options minimizes material waste and construction errors, contributing to cities building practices. Moreover, the ability to analyze important factors such as wind loads, pedestrian comfort and pollution dispersion provides a holistic view of the urban wind environment, offering information that is difficult to obtain using traditional methods. The robustness and efficiency of these models make them a valuable tool for engineering and building design, optimizing processes, reducing computational costs, and enabling sustainable and adaptive urban design. By combining CFD accuracy with ML speed, the proposed framework ensures resource-efficient, reliable, and practical solutions for complex wind environment analyses.

ML model limitations

Despite the advantages of the CFD-based ML model, its applicability is limited to the trained range of the dataset i.e. building dimensions and wind velocity. The model is expected to provide accurate predictions within the range and is generally not applicable to cases outside the range. This limitation highlights the need for careful consideration of the model's scope in practical applications.

Prospects for Future Research

The approach developed in the present work can be readily applied to more complicated cases. The development of CFD-based ML models offers promising opportunities for rapid and accurate

flow assessments around buildings, with several areas warranting further exploration. Expanding the model to account for more complex building geometries, including varied roof slopes, windows, and balconies, and increasing the number of buildings in the analysis could significantly enhance its applicability to real-world structures. Investigating the model's performance across a broader range of wind directions and weather conditions would provide additional insights into its potential applications for urban design. Future models could also incorporate complex atmospheric conditions, such as variable temperature gradients and thermal effects, to better simulate pollutant dispersion and pedestrian comfort in urban environments. Additionally, the current study focused validation primarily on the leeward side of the building. Expanding validation points to include critical areas, such as rooftops and windward sides, could enhance the reliability and accuracy of the model in capturing complex wind flow behaviors across various building regions. While this study provides a robust framework for simulating wind effects induced by buildings, the integration of additional environmental factors will further strengthen its utility for practical scenarios.

Finally, the potential of such hybrid models extends beyond urban wind studies and urban design. Future applications could include aerodynamic testing of vehicles, water pollution dispersion analysis, renewable energy assessments, and thermal comfort studies in outdoor urban spaces. This study underscores the value of CFD-based ML models as effective tools for accurate and efficient flow assessment across diverse engineering and environmental contexts.

5. Conclusions

This study presents an advanced integration of CFD and ML to accelerate and enhance the analysis of urban wind environments, including wind loads, pedestrian comfort, and pollution dispersion. By utilizing a dataset of 300 cases (100 buildings and three wind speeds) generated through CFD simulations, the Random Forest ML model achieved predictive accuracies ranging from 88% to 96%. The ML model significantly reduced computational time, delivering results in an

average of 96 milliseconds per case, compared to the 3 – 4 hours per case required by traditional CFD simulations.

The CFD-based ML approach not only significantly reduces computational costs but also provides a practical tool for architects, urban planners, and policymakers to evaluate design options rapidly and adaptively. By enabling quick and accurate predictions of wind flow and pollutant dispersion, this framework supports real-time urban planning, optimizing pedestrian comfort, and air quality management in dense environments. This capability is particularly crucial for sustainable city development and informed decision-making.

In addition to its methodological contributions, this study demonstrates the effectiveness of multi-output regression in CFD-based urban wind analysis, whereas previous works have primarily relied on single-output regression models or neural networks (e.g., DNNs and CNNs) for wind environment prediction. Furthermore, unlike some approaches that employ image-based feature extraction to obtain training data, this study directly utilizes numerical data from CFD mesh nodes, providing a more detailed and interpretable representation of the urban wind environment.

While single-output models are limited in their ability to predict multiple interdependent variables simultaneously, and neural networks often require large datasets and substantial computational resources, image-based extraction methods may introduce inaccuracies due to resolution constraints. In contrast, multi-output regression model effectively predicts multiple key parameters—wind velocity, turbulence intensity, and CO₂ mass fraction—while maintaining both high accuracy and computational efficiency. These findings contribute to the growing body of knowledge on CFD-based ML modeling, underscoring its potential as a robust and scalable method for simulating complex flows in urban environments.

Future research should expand the model’s applicability by incorporating more complex building groups, diverse architectural forms, wind directions, and atmospheric conditions such as temperature gradients, thereby enhancing its accuracy and versatility. Incorporating higher-fidelity turbulence models, such as like LES, could improve predictions, while larger and more diverse training datasets

could further enhance performance. Furthermore, integrating this methodology into interactive urban design platforms could provide urban planners with an accessible and computationally efficient tool for evaluating multiple design scenarios.

Potential future applications extend beyond urban wind comfort analysis. This approach can be adapted for renewable energy assessments, optimizing wind turbine placements in urban environments. Additionally, expanding the model to include pollutant dispersion from traffic emissions or industrial sources could enhance air quality management strategies. Finally, the application of this method to aerodynamic analysis of building clusters and urban heat island mitigation measures will aid sustainable urban planning.

By addressing these research directions, the proposed CFD-based ML model can evolve into a widely applicable tool for urban wind analysis, ultimately contributing to the creation of healthier, more sustainable, and climate-resilient cities.

Acknowledgments: This research was financially supported by a PhD scholarship from the Ministry of Science and Higher Education of Kazakhstan, as well as by the Erasmus Study Mobility Program. We would like to express our sincere gratitude to Rakhima Atabekova for her insightful comments and unwavering support throughout this study. We also extend our special thanks to Leonard Miraziz, IT specialist at Middlesex University, UK, for his technical contributions in enhancing computing power, which were crucial for the successful completion of these computational simulations, as well as for his invaluable assistance and support.

References

- [1] E. Nault, T. Jusselme, S. Aguacil, and M. Andersen, “Strategic environmental urban planning - A contextual approach for defining performance goals and informing decision-making,” *Build. Environ.*, vol. 168, p. 106448, Jan. 2020, doi: 10.1016/J.BUILDENV.2019.106448.
- [2] X. Zheng, H. Montazeri, and B. Blocken, “CFD simulations of wind flow and mean surface

- pressure for buildings with balconies: Comparison of RANS and LES,” *Build. Environ.*, vol. 173, p. 106747, Apr. 2020, doi: 10.1016/J.BUILDENV.2020.106747.
- [3] C. Baker, “Risk analysis of pedestrian and vehicle safety in windy environments,” *J. Wind Eng. Ind. Aerodyn.*, vol. 147, pp. 283–290, Dec. 2015, doi: 10.1016/J.JWEIA.2015.10.001.
- [4] J. Badach, W. Wojnowski, and J. Gębicki, “Spatial aspects of urban air quality management: Estimating the impact of micro-scale urban form on pollution dispersion,” *Comput. Environ. Urban Syst.*, vol. 99, p. 101890, Jan. 2023, doi: 10.1016/J.COMPENVURBSYS.2022.101890.
- [5] Y. Ishida, A. Yoshida, Y. Yamane, and A. Mochida, “Impact of a single high-rise building on the wind pressure acting on the surrounding low-rise buildings,” *J. Wind Eng. Ind. Aerodyn.*, vol. 250, p. 105742, Jul. 2024, doi: 10.1016/J.JWEIA.2024.105742.
- [6] P. J. Jones, D. Alexander, and J. Burnett, “Pedestrian Wind Environment Around High-Rise Residential Buildings in Hong Kong,” <http://dx.doi.org/10.1177/1420326X04045685>, vol. 13, no. 4, pp. 259–269, Aug. 2004, doi: 10.1177/1420326X04045685.
- [7] L. T. V and P. A. D, “The effects of wind on people in the vicinity of buildings.,” *Int Conf Wind Eff Build Struct*, vol. 4th, pp. 605–622, 1977, Accessed: Nov. 22, 2024. [Online]. Available: https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=201002029678401952.
- [8] B. Blocken and J. Carmeliet, “Pedestrian Wind Environment around Buildings: Literature Review and Practical Examples,” <http://dx.doi.org/10.1177/1097196304044396>, vol. 28, no. 2, pp. 107–159, Oct. 2004, doi: 10.1177/1097196304044396.
- [9] M. Li, G. Wang, X. Yang, and F. Yang, “Loss assessment of wind-induced damage for residential buildings groups based on engineering vulnerability,” *J. Build. Eng.*, vol. 42, p. 102435, Oct. 2021, doi: 10.1016/J.JOBE.2021.102435.
- [10] M. Xiong, B. Chen, H. Zhang, and Y. Qian, “Study on Accuracy of CFD Simulations of Wind Environment around High-Rise Buildings: A Comparative Study of k-ε Turbulence Models Based on Polyhedral Meshes and Wind Tunnel Experiments,” *Appl. Sci.* 2022, Vol.

12, Page 7105, vol. 12, no. 14, p. 7105, Jul. 2022, doi: 10.3390/APP12147105.

- [11] A. Kubilay, A. Rubin, D. Derome, and J. Carmeliet, “Wind-comfort assessment in cities undergoing densification with high-rise buildings remediated by urban trees,” *J. Wind Eng. Ind. Aerodyn.*, vol. 249, no. April, p. 105721, 2024, doi: 10.1016/j.jweia.2024.105721.
- [12] Y. Tamura, “Wind-Induced Damage To Buildings and Disaster,” *Seventh Asia-Pacific Conf. Wind Eng.*, 2009.
- [13] Y. Jing, H. Y. Zhong, W. W. Wang, Y. He, F. Y. Zhao, and Y. Li, “Quantitative city ventilation evaluation for urban canopy under heat island circulation without geostrophic winds: Multi-scale CFD model and parametric investigations,” *Build. Environ.*, vol. 196, p. 107793, Jun. 2021, doi: 10.1016/J.BUILDENV.2021.107793.
- [14] N. Abdollahzadeh and N. Bitoria, “Urban microclimate and energy consumption: A multi-objective parametric urban design approach for dense subtropical cities,” *Front. Archit. Res.*, vol. 11, no. 3, pp. 453–465, Jun. 2022, doi: 10.1016/J.FOAR.2022.02.001.
- [15] M. Amado and F. Poggi, “Solar Urban Planning: A Parametric Approach,” *Energy Procedia*, vol. 48, pp. 1539–1548, Jan. 2014, doi: 10.1016/J.EGYPRO.2014.02.174.
- [16] Y. Tominaga and M. Shirzadi, “Wind tunnel measurement of three-dimensional turbulent flow structures around a building group: Impact of high-rise buildings on pedestrian wind environment,” *Build. Environ.*, vol. 206, no. July, p. 108389, 2021, doi: 10.1016/j.buildenv.2021.108389.
- [17] B. Blocken, W. D. Janssen, and T. van 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. Softw.*, vol. 30, pp. 15–34, Apr. 2012, doi: 10.1016/J.ENVSOFT.2011.11.009.
- [18] “Visualizing the Trillion-Fold Increase in Computing Power.”
<https://www.visualcapitalist.com/visualizing-trillion-fold-increase-computing-power/>
(accessed Jun. 08, 2023).

- [19] M. S. Thordal, J. C. Bennetsen, S. Capra, A. K. Kragh, and H. H. H. Koss, "Towards a standard CFD setup for wind load assessment of high-rise buildings: Part 2 – Blind test of chamfered and rounded corner high-rise buildings," *J. Wind Eng. Ind. Aerodyn.*, vol. 205, p. 104282, Oct. 2020, doi: 10.1016/J.JWEIA.2020.104282.
- [20] Z. Kaseb, M. Hafezi, M. Tahbaz, and S. Delfani, "A framework for pedestrian-level wind conditions improvement in urban areas: CFD simulation and optimization," *Build. Environ.*, vol. 184, p. 107191, Oct. 2020, doi: 10.1016/J.BUILDENV.2020.107191.
- [21] Y. Li and L. Chen, "Study on the influence of voids on high-rise building on the wind environment," *Build. Simul.*, vol. 13, no. 2, pp. 419–438, 2020, doi: 10.1007/s12273-019-0584-7.
- [22] J. H. Ferziger, M. (Milovan) Perić, and R. L. Street, "Computational methods for fluid dynamics," p. 596.
- [23] M. Ali, "Computational Fluid Dynamics Computational Fluid Dynamics CSE-801," *Fluid Mech.*, pp. 293–327, 2018, Accessed: Aug. 02, 2023. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/B978012405935100006X>.
- [24] T. Lauriks *et al.*, "Application of Improved CFD Modeling for Prediction and Mitigation of Traffic-Related Air Pollution Hotspots in a Realistic Urban Street," *Atmos. Environ.*, vol. 246, p. 118127, Feb. 2021, doi: 10.1016/J.ATMOSENV.2020.118127.
- [25] T. Grylls, C. M. A. Le Cornec, P. Salizzoni, L. Soulhac, M. E. J. Stettler, and M. van Reeuwijk, "Evaluation of an operational air quality model using large-eddy simulation," *Atmos. Environ. X*, vol. 3, p. 100041, Jul. 2019, doi: 10.1016/J.AEAOA.2019.100041.
- [26] R. Yoshie *et al.*, "Cooperative project for CFD prediction of pedestrian wind environment in the Architectural Institute of Japan," *J. Wind Eng. Ind. Aerodyn.*, vol. 95, no. 9–11, pp. 1551–1578, Oct. 2007, doi: 10.1016/J.JWEIA.2007.02.023.
- [27] F. Xing, D. Mohotti, and K. Chauhan, "Study on localised wind pressure development in gable roof buildings having different roof pitches with experiments, RANS and LES

simulation models,” *Build. Environ.*, vol. 143, pp. 240–257, Oct. 2018, doi: 10.1016/J.BUILDENV.2018.07.026.

- [28] M. Lateb, C. Masson, T. Stathopoulos, and C. Bédard, “Comparison of various types of k-epsilon models for pollutant dispersion around a two-building configuration,” *J. Wind Eng. Ind. Aerodyn.*, vol. 115, pp. 9–21, 2013, [Online]. Available: <http://dx.doi.org/10.1016/j.jweia.2013.01.001>.
- [29] F. Toja-Silva, C. Peralta, O. Lopez-Garcia, J. Navarro, and I. Cruz, “Roof region dependent wind potential assessment with different RANS turbulence models,” *J. Wind Eng. Ind. Aerodyn.*, vol. 142, pp. 258–271, Jul. 2015, doi: 10.1016/J.JWEIA.2015.04.012.
- [30] M. Paris, F. Dubois, S. Bosc, and P. Devillers, “Integrating Wind Flow Analysis in Early Urban Design: Guidelines for Practitioners,” *J. Contemp. Urban Aff.*, vol. 7, no. 2, pp. 194–211, 2023, doi: 10.25034/ijcua.2023.v7n2-12.
- [31] J. P. Huijing, R. P. Dwight, and M. Schmelzer, “Data-driven RANS closures for three-dimensional flows around bluff bodies,” *Comput. Fluids*, vol. 225, p. 104997, 2021, doi: 10.1016/j.compfluid.2021.104997.
- [32] G. Merin Abbas and I. Gürsel Dino, “A parametric design method for CFD-supported wind-driven ventilation,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 609, no. 3, p. 032010, Sep. 2019, doi: 10.1088/1757-899X/609/3/032010.
- [33] Y. Qiu, Y. He, M. Li, and X. Zhu, “A Generalization of Building Clusters in an Urban Wind Field Simulated by CFD,” *Atmos. 2024, Vol. 15, Page 9*, vol. 15, no. 1, p. 9, Dec. 2023, doi: 10.3390/ATMOS15010009.
- [34] D. Panchigar, K. Kar, S. Shukla, R. M. Mathew, U. Chadha, and S. K. Selvaraj, “Machine learning-based CFD simulations: a review, models, open threats, and future tactics,” *Neural Comput. Appl.*, vol. 34, no. 24, pp. 21677–21700, Dec. 2022, doi: 10.1007/S00521-022-07838-6/METRICS.
- [35] S. C. Koumetio Tekouabou, E. B. Diop, R. Azmi, R. Jaligot, and J. Chenal, “Reviewing the

application of machine learning methods to model urban form indicators in planning decision support systems: Potential, issues and challenges,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5943–5967, Sep. 2022, doi: 10.1016/J.JKSUCI.2021.08.007.

- [36] H. Wang, J. Yang, G. Chen, C. Ren, and J. Zhang, “Machine learning applications on air temperature prediction in the urban canopy layer: A critical review of 2011–2022,” *Urban Clim.*, vol. 49, p. 101499, May 2023, doi: 10.1016/J.UCLIM.2023.101499.
- [37] A. Król, “The Application of the Artificial Intelligence Methods for Planning of the Development of the Transportation Network,” *Transp. Res. Procedia*, vol. 14, pp. 4532–4541, Jan. 2016, doi: 10.1016/J.TRPRO.2016.05.376.
- [38] T. Van Quang, D. T. Doan, and G. Y. Yun, “Recent advances and effectiveness of machine learning models for fluid dynamics in the built environment,” *Int. J. Model. Simul.*, Jun. 2024, doi: 10.1080/02286203.2024.2371682.
- [39] Y. Casali, N. Y. Aydin, and T. Comes, “Machine learning for spatial analyses in urban areas: a scoping review,” *Sustain. Cities Soc.*, vol. 85, p. 104050, Oct. 2022, doi: 10.1016/J.SCS.2022.104050.
- [40] K. Wijesooriya, D. Mohotti, C. K. Lee, and P. Mendis, “A technical review of computational fluid dynamics (CFD) applications on wind design of tall buildings and structures: Past, present and future,” *J. Build. Eng.*, vol. 74, p. 106828, Sep. 2023, doi: 10.1016/J.JOBE.2023.106828.
- [41] G. Calzolari and W. Liu, “Deep learning to replace, improve, or aid CFD analysis in built environment applications: A review,” *Build. Environ.*, vol. 206, p. 108315, Dec. 2021, doi: 10.1016/J.BUILDENV.2021.108315.
- [42] A. Refaie Ali, R. Mahmood, A. Asghar, A. H. Majeed, and M. H. Behiry, “AI-based predictive approach via FFB propagation in a driven-cavity of Ostwald de-Waele fluid using CFD-ANN and Levenberg–Marquardt,” *Sci. Rep.*, vol. 14, no. 1, pp. 1–13, 2024, doi: 10.1038/s41598-024-60401-2.

- [43] T. Zhang, B. Dey, K. Veeraraghavan, H. Kulkarni, and A. Chakraborty, “Demystifying the Data Need of ML-surrogates for CFD Simulations,” May 2022, Accessed: Nov. 22, 2024. [Online]. Available: <https://arxiv.org/abs/2205.08355v1>.
- [44] K. Du and B. Chen, “A hybrid semi-supervised regression based machine learning method for predicting peak wind loads on a group of buildings,” *Eng. Struct.*, vol. 275, p. 115245, Jan. 2023, doi: 10.1016/J.ENGSTRUCT.2022.115245.
- [45] M. R. Kaloop, A. Bardhan, P. Samui, J. W. Hu, and M. Elsharawy, “Comparative study on deep and machine learning approaches for predicting wind pressures on tall buildings,” *Alexandria Eng. J.*, vol. 111, pp. 610–627, Jan. 2025, doi: 10.1016/J.AEJ.2024.10.047.
- [46] A. M. Aly and J. Clarke, “Wind design of solar panels for resilient and green communities: CFD with machine learning,” *Sustain. Cities Soc.*, vol. 94, p. 104529, Jul. 2023, doi: 10.1016/J.SCS.2023.104529.
- [47] Y. He *et al.*, “Hybrid framework for rapid evaluation of wind environment around buildings through parametric design, CFD simulation, image processing and machine learning,” *Sustain. Cities Soc.*, vol. 73, p. 103092, Oct. 2021, doi: 10.1016/J.SCS.2021.103092.
- [48] J. Wehrle, C. Jung, M. Giometto, A. Christen, and D. Schindler, “Introducing new morphometric parameters to improve urban canopy air flow modeling: A CFD to machine-learning study in real urban environments,” *Urban Clim.*, vol. 58, p. 102173, 2024, doi: 10.1016/j.uclim.2024.102173.
- [49] R. Zhao, S. Liu, J. Liu, N. Jiang, and Q. Chen, “A two-stage CFD-GNN approach for efficient steady-state prediction of urban airflow and airborne contaminant dispersion,” *Sustain. Cities Soc.*, vol. 112, p. 105607, Oct. 2024, doi: 10.1016/J.SCS.2024.105607.
- [50] M. Mendil, S. Leirens, P. Armand, and C. Duchenne, “Hazardous atmospheric dispersion in urban areas: A Deep Learning approach for emergency pollution forecast,” *Environ. Model. Softw.*, vol. 152, p. 105387, Jun. 2022, doi: 10.1016/J.ENVSOFT.2022.105387.
- [51] V. Yadav, A. K. Yadav, V. Singh, and T. Singh, “Artificial neural network an innovative

- approach in air pollutant prediction for environmental applications: A review,” *Results Eng.*, vol. 22, no. February, p. 102305, 2024, doi: 10.1016/j.rineng.2024.102305.
- [52] X. Lin *et al.*, “CFD- and BPNN- based investigation and prediction of air pollutant dispersion in urban environment,” *Sustain. Cities Soc.*, vol. 100, p. 105029, Jan. 2024, doi: 10.1016/J.SCS.2023.105029.
- [53] J. Lee, C. A. M. Barquilla, K. Park, and A. Hong, “Urban form and seasonal PM2.5 dynamics: Enhancing air quality prediction using interpretable machine learning and IoT sensor data,” *Sustain. Cities Soc.*, vol. 117, p. 105976, Dec. 2024, doi: 10.1016/J.SCS.2024.105976.
- [54] B. E. Launder and D. B. Spalding, “The numerical computation of turbulent flows,” *Comput. Methods Appl. Mech. Eng.*, vol. 3, no. 2, pp. 269–289, Mar. 1974, doi: 10.1016/0045-7825(74)90029-2.
- [55] V. Yakhot, S. A. Orszag, S. Thangam, T. B. Gatski, and C. G. Speziale, “Development of turbulence models for shear flows by a double expansion technique,” *Phys. Fluids A Fluid Dyn.*, vol. 4, no. 7, pp. 1510–1520, 1992, Accessed: Jun. 08, 2023. [Online]. Available: https://www.academia.edu/25466983/Development_of_turbulence_models_for_shear_flows_by_a_double_expansion_technique.
- [56] T.-H. Shih, W. W. Liou, A. Shabbir, Z. Yang, and J. Zhu, “A New K-epsilon Eddy Viscosity Model for High Reynolds Number Turbulent Flows,” *Computers Fluids*, vol. 24, no. 3, pp. 227–238, 1995.
- [57] ANSYS FLUENT, “ANSYS Fluent Theory Guide,” *ANSYS Inc., USA*, vol. 15317, no. November, pp. 1–759, 2013, [Online]. Available: [http://www.pmt.usp.br/ACADEMIC/martoran/NotasModelosGrad/ANSYS Fluent Theory Guide 15.pdf](http://www.pmt.usp.br/ACADEMIC/martoran/NotasModelosGrad/ANSYS_Fluent_Theory_Guide_15.pdf).
- [58] Y. Tominaga and T. Stathopoulos, “CFD simulations of near-field pollutant dispersion with different plume buoyancies,” *Build. Environ.*, vol. 131, pp. 128–139, Mar. 2018, doi:

10.1016/j.buildenv.2018.01.008.

- [59] S. V. Patankar, “Numerical Heat Transfer and Fluid Flow,” Oct. 2018, doi: 10.1201/9781482234213.
- [60] Y. Tominaga and T. Stathopoulos, “CFD simulation of near-field pollutant dispersion in the urban environment: A review of current modeling techniques,” 2013, doi: 10.1016/j.atmosenv.2013.07.028.
- [61] J. C. Chang and S. R. Hanna, “Air quality model performance evaluation,” *Meteorol Atmos Phys*, vol. 87, pp. 167–196, 2004, doi: 10.1007/s00703-003-0070-7.
- [62] A. Issakhov and A. R. Baitureyeva, “Numerical modelling of a passive scalar transport from thermal power plants to air environment,” *Adv. Mech. Eng.*, vol. 10, no. 10, Oct. 2018, doi: 10.1177/1687814018799544/ASSET/IMAGES/LARGE/10.1177_1687814018799544-FIG12.JPEG.
- [63] A. A. Issakhov and A. R. Baitureyeva, “Modeling of a passive scalar transport from thermal power plants to atmospheric boundary layer,” *Int. J. Environ. Sci. Technol.*, vol. 16, no. 8, pp. 4375–4392, Aug. 2019, doi: 10.1007/S13762-019-02273-Y.
- [64] A. Issakhov, T. Yang, and A. Baitureyeva, “CFD simulation of pollution dispersion from thermal power plants in the atmosphere,” *Int. J. Math. Phys.*, vol. 10, no. 1, pp. 56–65, Jun. 2019, doi: 10.26577/IJMPH-2019-II-8.
- [65] A. Issakhov and A. Baitureyeva, “Numerical simulation of a pollutant transport in lower atmosphere layer from thermal power plants,” *WSEAS Trans. Fluid Mech.*, vol. 12, pp. 16–32, 2017.
- [66] A. Issakhov and A. Baitureyeva, “Numerical simulation of a passive scalar transport from thermal power plants,” *AIP Conf. Proc.*, vol. 1836, Jun. 2017, doi: 10.1063/1.4981959.
- [67] A. Issakhov and A. Baitureyeva, “Numerical study of a passive scalar transport in lower atmosphere layer from thermal power plants,” *AIP Conf. Proc.*, vol. 1978, no. 1, Jul. 2018, doi: 10.1063/1.5044122/772308.

- [68] A. Issakhov and A. Baitureyeva, “Numerical study of a passive scalar transport from thermal power plants to air environment,” *Commun. Comput. Inf. Sci.*, vol. 998, pp. 114–124, 2019, doi: 10.1007/978-3-030-12203-4_11.
- [69] J. Franke, A. Hellsten, K. H. Schlünzen, and B. Carissimo, “Best practice guideline for the CFD simulation of flows in the urban environment - a summary.” Cambridge Environmental Research Consultants, 2007, Accessed: Jun. 08, 2023. [Online]. Available: <https://research.aalto.fi/en/publications/best-practice-guideline-for-the-cfd-simulation-of-flows-in-the-ur>.
- [70] P. J. Richards and R. P. Hoxey, “Appropriate boundary conditions for computational wind engineering models using the k- ϵ turbulence model,” *J. Wind Eng. Ind. Aerodyn.*, vol. 46–47, no. C, pp. 145–153, Aug. 1993, doi: 10.1016/0167-6105(93)90124-7.
- [71] A. G. Davenport, C. S. B. Grimmond, T. R. Oke, and J. Wieringa, “Estimating the roughness of cities and sheltered country,” *15th Conf. Probab. Stat. Atmos. Sci. Conf. Appl. Climatol. Asheville, NC, Am. Meteorol. Soc.*, no. August, pp. 96–99, 2000.
- [72] “Simulated historical climate & weather data for Almaty - meteoblue.” https://www.meteoblue.com/en/weather/historyclimate/climatemodelled/almaty_kazakhstan_1526395 (accessed Jun. 08, 2023).
- [73] T. Beer, “Environmental Oceanography,” *Environ. Oceanogr.*, Dec. 2017, doi: 10.1201/9780203756980.
- [74] Y. Takano and P. Moonen, “On the influence of roof shape on flow and dispersion in an urban street canyon,” *J. Wind Eng. Ind. Aerodyn.*, vol. 123, pp. 107–120, Dec. 2013, doi: 10.1016/J.JWEIA.2013.10.006.
- [75] B. Blocken, T. Stathopoulos, and J. Carmeliet, “CFD simulation of the atmospheric boundary layer: wall function problems,” *Atmos. Environ.*, vol. 41, no. 2, pp. 238–252, Jan. 2007, doi: 10.1016/J.ATMOSENV.2006.08.019.
- [76] L. F. Richardson, “On the approximate arithmetical solution by finite differences of physical

problems involving differential equations, with an application to the stresses in a masonry dam,” *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character*, vol. 83, no. 563, pp. 335–336, Mar. 1910, doi: 10.1098/RSPA.1910.0020.

- [77] P. J. Roache, “Perspective: A Method for Uniform Reporting of Grid Refinement Studies,” *J. Fluids Eng.*, vol. 116, no. 3, pp. 405–413, Sep. 1994, doi: 10.1115/1.2910291.
- [78] S. Vinchurkar and P. W. Longest, “Evaluation of hexahedral, prismatic and hybrid mesh styles for simulating respiratory aerosol dynamics,” *Comput. Fluids*, vol. 37, no. 3, pp. 317–331, Mar. 2008, doi: 10.1016/J.COMPFLUID.2007.05.001.
- [79] I. Forys, “Machine learning in house price analysis: regression models versus neural networks,” *Procedia Comput. Sci.*, vol. 207, pp. 435–445, Jan. 2022, doi: 10.1016/J.PROCS.2022.09.078.
- [80] A. C. Müller and S. Guido, “Introduction to Machine Learning with Python A GUIDE FOR DATA SCIENTISTS Introduction to Machine Learning with Python.”
- [81] A. Botchkarev, “Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology,” *Interdiscip. J. Information, Knowledge, Manag.*, vol. 14, pp. 45–76, 2018, doi: 10.28945/4184.
- [82] Y. Liu, Y. Wang, and J. Zhang, “New Machine Learning Algorithm: Random Forest,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7473 LNCS, pp. 246–252, 2012, doi: 10.1007/978-3-642-34062-8_32.
- [83] T. V. Lawson, “The wind content of the built environment,” *J. Wind Eng. Ind. Aerodyn.*, vol. 3, no. 2–3, pp. 93–105, Jan. 1978, doi: 10.1016/0167-6105(78)90002-8.