This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

Data-Driven Prediction of Wind Pressure on Low-Rise Buildings in Complex Heterogeneous Terrains

Lee-Sak An[1,2] and Sungmoon Jung[2]*

Abstract

This study presents a data-driven methodology for predicting the pressure coefficient statistics on the windward wall, roof, and leeward wall of low-rise buildings situated downwind of complex heterogeneous terrains. Two types of artificial neural network models were developed: the empirical parameter-based ANN (PANN) and the morphology-based ANN (MANN). Pressure data from wind tunnel tests on the Wind Engineering Research Field Laboratory (WERFL) building model (building height $H = 4$ m) in complex heterogeneous terrain were used to develop the ANN models. These models were evaluated against a non-linear fitted model to assess their predictive performances. PANN and MANN demonstrated superior performance in capturing the effects of terrain complexity on the mean ($C_{p,mean}$) and the root-mean-square ($C_{p,RMS}$) wind pressure coefficients for the windward wall, roof, and leeward wall. Optimal prediction was achieved with a terrain patch size of W×L=4×2, equating to a full-scale area of approximately 72 m × 23 m. This suggests that the morphology within approximately 100 m × 50 m ($25H \times 12.5H$) in front of a low-rise building has the greatest correlation with the wind pressure coefficient. Despite lower R^2 values for max $C_{p,RMS}$ on the leeward wall across all models, both PANN and MANN showed promising accuracy for the six outputs studied. Moreover, a global sensitivity analysis confirmed

---

[1] Korea Floating Infrastructure Research Center, Seoul National University, 1 Gwanakro, Gwanak-gu, Seoul, 08826, South Korea.

[2] Department of Civil and Environmental Engineering, FAMU-FSU College of Engineering, 2035 E Paul Dirac Dr, Tallahassee, Florida 32310, United States.

* Corresponding author (Email: sjung@eng.famu.fsu.edu)

the impact of terrain roughness and complexity on the prediction models particularly on max $C_{p,RMS}$, and underscored the dominance of effective roughness length $z_{0,eff}$ and the coefficient of variation of roughness length COV_{z_0} in influencing model outcomes.

Keywords

Low-rise building; Wind pressure coefficient; Complex heterogeneous terrain; Artificial Neural Network; Bayesian optimization

1. Introduction

Terrain configuration is a critical factor in introducing uncertainties in wind loads, as underscored in the Davenport's wind loading chain [1]. The influence of terrain roughness becomes particularly pronounced for low-rise buildings situated near the ground surface, as they are exposed to increased turbulence. Although the majority of current knowledge is confined to homogeneous (i.e., uniform) terrain, terrains in the real world are often complex and have abrupt changes in surface roughness. Especially, upstream terrain configurations within a short distance upwind of a site have a direct impact on wind loads on building envelopes [2].

Significant knowledge gaps still remain regarding the influence of the complex heterogeneous terrain on the pressure experienced by low-rise buildings. Only a few studies have discussed the effect of terrain complexity on wind loads. Yu et al. [3] conducted wind tunnel tests using two real city terrain models and proposed a minimum upstream patch length for wind tunnel testing. They experimentally revealed that mean velocity profiles in urban areas are influenced by an upstream patch length up to 750 m, and are not affected by the patch that exceeds 1250 m. Wang and Stathopoulos [2] emphasized the significance of local, small-scale roughness changes in affecting the variation of the wind speed profile above heterogeneous terrain. Kim et al. [4] investigated the effect of a large group of surrounding buildings on a typical low-rise building by measuring wind pressure. They observed that, although the mean pressure coefficient decreased, the peak pressure coefficient could increase due to the enhancement of the turbulence component. An et al. [5] conducted extensive wind tunnel testing to explore wind characteristics over complex heterogeneous terrains. They quantified the relationship between the variance of geometric morphology and wind characteristics, ultimately concluding that terrain complexity significantly increased turbulence intensity levels. Subsequently, An and Jung [6] investigated the wind pressure coefficients on the windward wall and roof and quantified the influence of terrain complexity on the pressure behaviors of low-rise buildings. Kim et al. [7] experimentally delved into the complex dynamics of upwind terrain transition from open country to suburban areas and its effects on wind pressures and forces on low-rise buildings. It is anticipated that pressure coefficients over complex heterogeneous terrains will differ from those over homogeneous terrains due to the substantial influence of turbulence properties in the approaching wind flow on the pressure field [8, 9]. However, it is still challenging to predict the highly variable wind pressure on low-rise buildings over complex heterogeneous terrains due to the lack of field-measured and experimental data. Until recently, there has been no field-measured or experimental data on the variability of wind pressure that can be caused by the wide variety of terrain that exists in the real world.

Recent evidence suggests that artificial neural network (ANN) methods, as a data-driven approach, are particularly effective in addressing problems in wind engineering due to their robustness in solving multivariate and nonlinear regression problems. Numerous studies have demonstrated the efficacy of ANN methods for predicting wind pressure on building structures. Gavalda et al. [10] studied variable plan dimensions and roof slope in a set of parameters considered in earlier interpolation studies using ANN. Chen et al. [11] attempted predict wind pressure on low-rise buildings using an ANN approach. They proposed using ANN for the prediction of wind pressure time series. Bre et al. [12] adopted an ANN to predict the surface-average pressure coefficients for each wall and roof according to the building geometries and the wind directions. Fernández-Cabán et al. [13] applied an ANN to predict roof pressures on low-rise structures based on freestream turbulence conditions. The ANN model was trained and tested using a comprehensive dataset from a recent boundary layer wind tunnel pressure dataset for homogeneous terrain cases. Tian et al. [14] also applied an ANN for predicting mean and peak wind pressure coefficients on the surface of low-rise, gable roof buildings. They suggested that with a large enough database, the ANN-based method could significantly enhance knowledge yield and reduce experimental effort. Ding et al. [15] developed and optimized ANN models for predicting wind pressures on low-rise buildings using genetic algorithms and Bayesian optimization. They evaluated the influences of the hyperparameters, the number of data pairs, and the ANN structures on their performances. Lang et al. [16] proposed and verified the performance of an improved random forest algorithm for predicting the mean and fluctuating wind pressure coefficients of high-rise buildings. Although such studies have used data-driven approaches for predicting wind pressure on buildings, these studies have been limited for homogeneous terrains due to the lack of wind pressure dataset over complex heterogeneous terrains. Neither ANN nor any other data-driven approaches can be found in the literature to predict the statistics of C_P on buildings over complex heterogeneous terrains, and understanding the features that can be used to assess the variability of wind pressure according to the degree of terrain complexity is still insufficient. Due to cost and time constraints, the full-scale or wind tunnel experiments commonly

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

entail a limited number of approach flow conditions, and such limitations of the dataset prevented the application of data-driven approaches.

This study developed ANN models for predicting the peak values of mean pressure coefficients ($C_{p,mean}$) and root-mean-square pressure coefficients ($C_{p,RMS}$) on low-rise buildings. The recently released wind tunnel testing dataset over complex heterogeneous terrains was used for training the ANN models. As the very first study on the data-driven approach for predicting wind pressure over complex heterogeneous terrains, the best input features that can represent the terrain complexity level were investigated. The two different types of input features were applied to train the ANN models: empirical parameters (indirect information) and morphology (direct information). Sensitivity analysis was performed on each model to analyze the degree of influence on the wind pressure coefficient. Moreover, we proposed the most appropriate patch size for predicting the wind pressure coefficient by comparing the prediction performance of ANN models using different patch sizes of the terrain. The ANN models, the proposed input features, and the patch size will be valuable preliminary research for future research on wind pressure in complex heterogeneous terrains.

2. Methodology

2.1. Wind Tunnel Test Dataset

In this section, we provide a brief overview of the wind tunnel test dataset used for developing the ANN models. Alinejad et al. [17] offers comprehensive details about the test setup. For further details on the site selection, reproducing heterogeneous terrains from the real sites, and an in-depth investigation into the wind pressure coefficients refer to An et al. [5], An and Jung [6], and Alinejad et al. [18].

The wind tunnel testing was carried out at the Natural Hazard Engineering Research Infrastructure (NHERI) experimental facility situated at the University of Florida [19]. Fig. 1 illustrates the schematic layout of the wind tunnel facility, which is an open circuit tunnel with dimensions of 6 m (width) × 3 m (height) × 38 m (length). The tunnel inlet incorporates eight vane axial fans, each driven by a 56-kW electric motor. The flow generated by these fans is conditioned by honeycombs positioned approximately 3 m downwind from the fan bank.

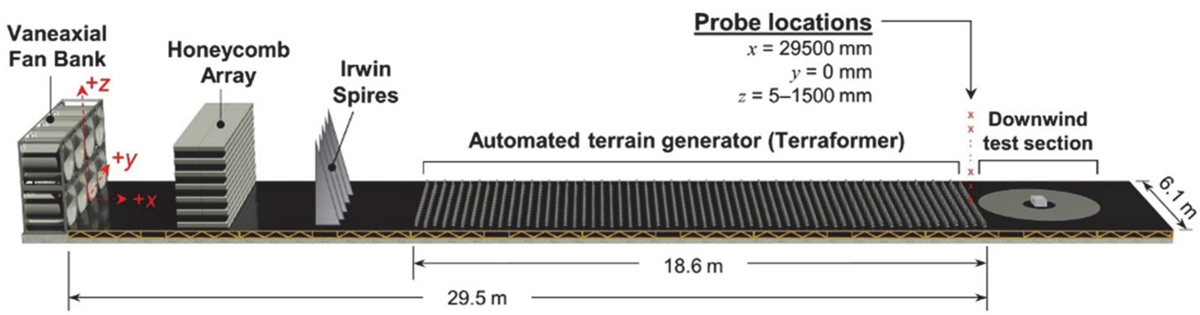


Fig. 1. Schematic plan of the wind tunnel facility at the University of Florida [20].

This facility houses a fully automated terrain simulator named the "Terraformer." This state-of-the-art technology enables the swift and precise simulation of terrain, addressing the time-consuming and labor-intensive challenges associated with wind tunnel testing. The Terraformer consists of an 18 × 62 array of computer-controlled roughness blocks (total 1116 elements) in a staggered layout, covering a patch size of 6.1 m × 18.6 m. Each roughness element is equipped with an actuator, allowing for independent height adjustments. These elements have a plan dimension of 100 mm × 50 mm and adjustable heights ranging from 0 to 160 mm. The reconfiguration of all 1116 elements typically takes less than 60 s, making the Terraformer an efficient tool for simulating a wide range of homogeneous and heterogeneous upwind terrains. Additionally, a turntable located at the end of the upwind patch enables the simulation of wind effects on structures at various wind incidence angles. Wind tunnel experiments varied the wind incident angle (α) to 0°, 15°, 30°, 45°, 60°, 75°, and 90°.

Previous research on low-rise structures in boundary layer wind tunnels has indicated that accurately replicating full-scale turbulence characteristics (such as integral length scale) at or near the height of the building model is crucial for precisely quantifying extreme aerodynamic loads, especially in regions of flow separation. As Stathopoulos [21] summarized, simulating only the lower region of the atmospheric surface layer (ASL) with larger model scales (such as 1:50 to 1:100) is an effective approach for addressing the length scale problem encountered in wind tunnel testing. Numerous previous wind tunnel tests for low-rise buildings were conducted using model scales within this range [22-25]. This test adopted a 1:50 scale, indicating the maximum vertical measurement height of 1500 mm in test scale corresponds to 75 m in full-scale representation. Similarly, the Terraformer simulates terrain of 930 m on a full-scale. This satisfies the upstream fetch considered significant (1 km) when assessing wind loads on lower buildings (building height<50 m) [26, 27]. The low-rise building model has dimensions of 274 mm × 182 mm × 80 mm in testing scale (13.7 m × 9.1 m × 4 m in full-scale) with a 1/4:12 gable roof slope, mirroring the design of the Wind Engineering Research Field Laboratory (WERFL) building at Texas Tech University [28].

Pressure measurements were acquired using eight high-speed electronic scanning modules from Scanivalve ZOC33 [29]. Pressure taps are connected to the modules via 122 cm long urethane tubing, and the sampling frequency was set at 625 Hz. Adjustments were made to minimize tubing effects on pressure measurements, reducing distortion on amplitude and phase shift [30]. Pressure data were recorded based on the time series.

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

Fig. 2 provides a visual representation of the pressure tap layout on the low-rise building model. The building model was outfitted with a total of 206 pressure taps, comprising 92 roof taps and 114 wall taps. The tap positions adhered to the layout used in the WERFL model of the NIST aerodynamic database [31].

Our aim was to predict peak values of wind pressure coefficient statistics. The pressure coefficient at a point of interest, denoted as C_p, is defined as the ratio between the measured building surface gauge pressure and the roof-height dynamic pressure, expressed by the formula:

$$C_p(t) = \frac{p(t) - p_0}{0.5\rho U_H^2} \tag{1}$$

Here, U_H represents the wind speed at the eave height of the low-rise building (4 m), and ρ denotes the air density. The term $p(t) - p_0$ signifies the net wind pressure at the point of interest, with p_0 referring to the reference pressure. The representative tap line was selected to capture the peak of C_p statistics, particularly peak mean ($C_{p,mean}$) and maximum root-mean-square ($C_{p,RMS}$) of C_p. This tap line, comprising the series of taps closest to the center of the low-rise building in the perpendicular direction of the ridge, has been consistently employed in previous studies to examine flow separation and reattachment behavior on the building surfaces [8, 31]. Fig. 3 showcases the definition of peak $C_{p,mean}$ and max $C_{p,RMS}$.

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

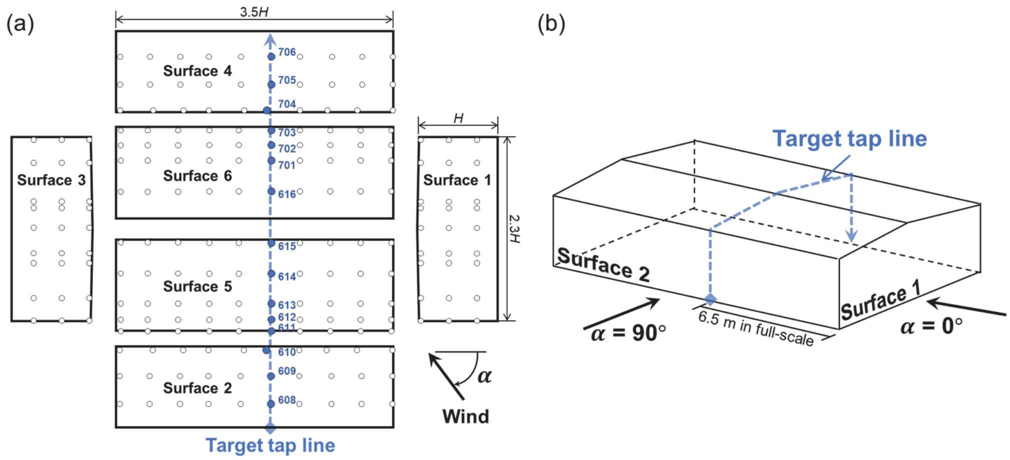


Fig. 2. The low-rise building model and tap information: (a) Plan view, and (b) 3D view.

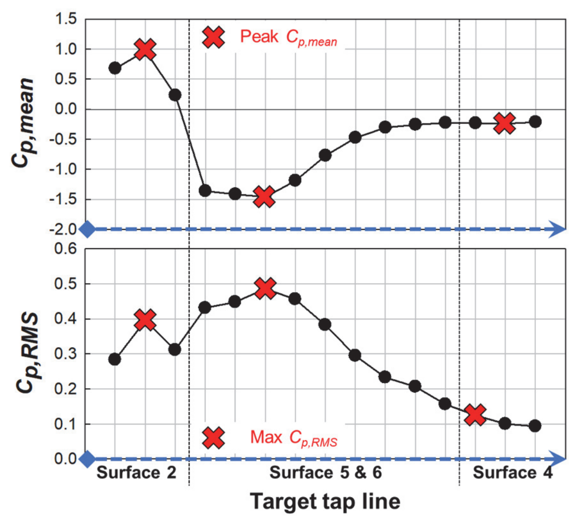


Fig. 3. Definition of peak $C_{p,mean}$ and max $C_{p,RMS}$.

Complex heterogeneous terrain configurations drawn from real terrains were compiled for wind tunnel testing. The primary data source was the National Land Cover Database (NLCD) [32] provided by the US Geological Survey. A total of 529 sites from 32 US states prone to hurricanes were selected. The k-means algorithm [33] was used in the 2D space defined by the mean and standard deviation to select representative terrains with distinct stochastic properties of local roughness length, leading to the identification and classification of 50 distinct clusters. Thus, the

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

50 representative terrains were conclusively selected from 529 sites in the US. In the wind tunnel, these roughness lengths were correlated with the corresponding block heights [34]. The details of producing heterogeneous terrains in the wind tunnel were described by Alinejad et al. [35]. Fig. 4 provides examples of the selected sites and their corresponding block height maps in the Terraformer, along with the simulated terrain morphology generated for site 8. Since wind tunnel experiments varied the wind incident angle (α) to seven cases, a total of 350 datasets (7 angles × 50 terrains) was provided for the development of the prediction model.

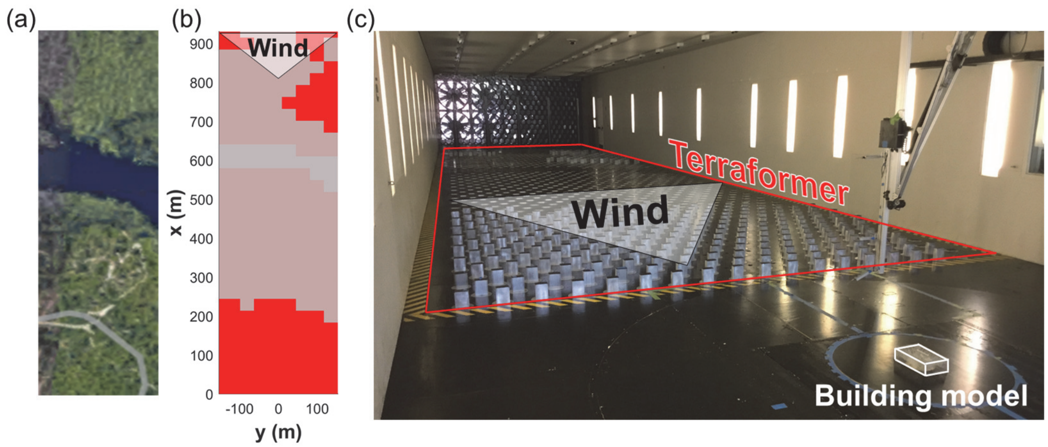


Fig. 4. Example of complex heterogeneous terrains (site 8): (a) Aerial view (from Google Earth); (b) Block height map; and (c) Actual photo in the wind tunnel.

2.2. Artificial Neural Network

Artificial Neural Network (ANN) operates by processing information through interconnected nodes in layers; input data is fed into the network, processed through one or more hidden layers where each node computes weighted sums of its inputs followed by an activation function, and finally produces an output through the output layer. This architecture enables the network to learn complex nonlinear relations between the input and output pairs by adjusting the weights during the training process. Fig. 5 shows the architecture of an ANN, which typically includes n-input nodes, an output node, and one or more hidden layers. ANN models for predicting peak $C_{p,mean}$ and max $C_{p,RMS}$ on the windward wall, roof, and leeward wall (total 6 models) were independently

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

developed. The objective function incorporated the mean squared error (MSE) loss function and the ridge (L2) penalty term, and backpropagation was used to adjust the weights. The loss function in an ANN measures the difference between the predicted output and the actual target value. It guides the optimization process by quantifying how well or poorly the model performs, with the purpose of minimizing this error to improve the model's accuracy.

We developed two types of ANN models using different input features related to the level of terrain complexity. The first type is a parameter-based ANN model (PANN). The empirical parameters–effective roughness length $z_{0,eff}$ and the coefficient of variation for the z_0 values in the terrain COV_{z_0}–used as input features were determined based on the investigations from the author's previous studies [5, 6]. In our prior work, An et al. [5] investigated the influence of terrain complexity on the mean wind profile and turbulence intensity. They found that the $z_{0,eff}$, widely used in atmospheric surface layer modeling for moderately homogeneous terrain or smaller-scale inhomogeneity, was insufficient for complex heterogeneous terrain. Additional consideration of the morphological variation of the terrain was necessary. It was found that the $z_{0,eff}$ and the COV_{z_0} are dominant parameters affecting wind characteristics over complex heterogeneous terrains. Subsequently, An and Jung [6] found that wind characteristics influenced by terrain complexity affected the variability of wind pressure on low-rise buildings, even though the $z_{0,eff}$ was similar. Thus, PANN used three input features: $z_{0,eff}$, COV_{z_0}, and wind incident angle α.

$z_{0,eff}$ was calculated using a grid-squared average-based approach, utilizing the z_0 maps of the terrains [36, 37]. This approach relied on the linear approximation of the Rossby number similarity theory and derived the following formula [36]:

$$\ln\left(z_{0,eff}\right) = \langle \ln\,(z_0) \rangle + a\sigma^2_{ln(z_0)} \qquad (2)$$

Here, a represents the Rossby value, typically set to 0.09, and $\sigma^2_{ln(z_0)}$ indicates the variance within the area. The $\langle \rangle$ notation represents the area-weighted logarithmic average operation. COV_{z_0} can be calculated as the standard deviation/average of z_0 values in a given map. The z_0 maps can be attained by transforming the block height map, shown in Fig. 4 (b). The relationship between block height and z_0 is outlined in Appendix B of An et al. [5]. They performed a wind tunnel experiment by uniformly changing the block height and then used an anemometric approach [38] to estimate z_0 caused by each block height. By using the block height vs z_0 relationship obtained from the estimated results, a z_0 map of the given terrain can be obtained.

The second type is a morphology-based ANN model (MANN). Direct use of morphology in wind loading estimation was shown to be effective in a previous study [39]. Here, the values of the z_0 maps were directly used as input features, along with α. The number of input features equals the product of the number of x-direction blocks (L) and the number of y-direction blocks (W).

For both approaches, the model development was iteratively conducted, while changing patch size to determine the optimal patch size that showed the best prediction performance. As investigated in previous studies, the roughness of the terrain at a certain distance from the building does not have a significant effect on the wind pressure [2, 3]. Thus, considering information from a wider patch does not guarantee higher wind pressure prediction accuracy. In the case of PANN, $z_{0,eff}$ and COV_{z_0} change depending on the patch size being considered, so if the morphology of an excessively wide patch is considered, the correlation with wind pressure may decrease and prediction performance may deteriorate. Moreover, for MANN, patch size is a dominant factor affecting the effectiveness of the model training since the number of input features is the number of blocks in the considered patch sizes. By comparing the prediction performance with changing

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

the patch size, the best patch size that showed the greatest correlation with the characteristics of wind pressure was explored.

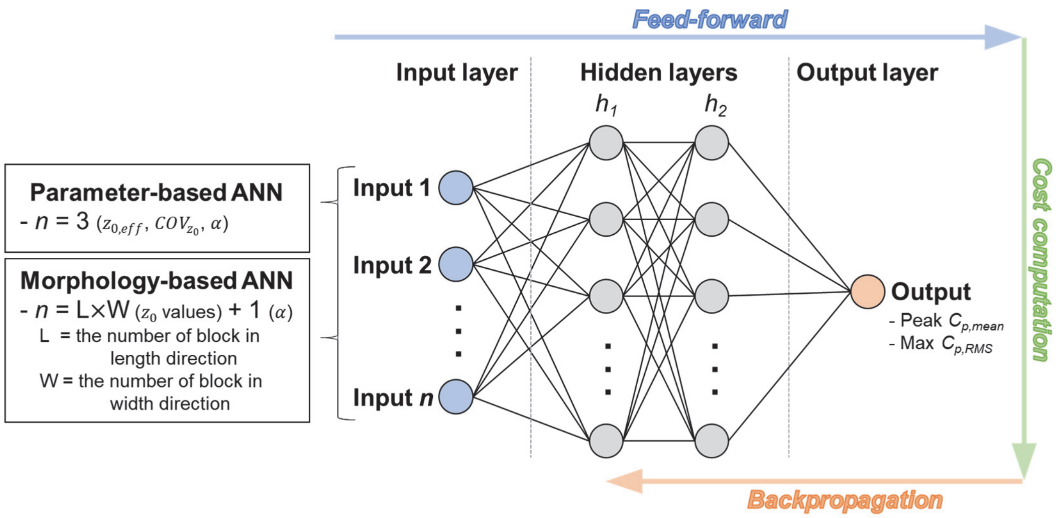


Fig. 5. Architecture of ANN and input features of PANN and MANN.

In developing ANN models, architecture optimization can significantly improve accuracy since a more efficient and carefully designed architecture achieves good generalization and avoids overfitting. Optimal network structure is mostly determined by the data nature rather than the sample size, suggesting a data-driven approach to choosing the ANN architecture [40]. To determine the optimal hyperparameters, Bayesian optimization (BO) is applied in this study. BO has been recognized as an excellent tool to find the global optimum with a minimum number of steps and has outperformed other state-of-the-art global optimization algorithms on some challenging optimization benchmarks [41]. The strategy of BO assumes the unknown objective function as a random function and places a prior over it, which captures beliefs about the behavior of the objective function. Ding et al. [15] reported that the BO-based neural network (BONN) was most efficient, saving 88-94% computational time compared with the traditional trial-and-error neural network. The effectiveness of BONN for wind pressure prediction has already been validated in previous studies [41, 42]. The number of layers and the number of nodes ranged from

1-2 and 1-50, respectively. The optimization options for the activation function were none, relu, tanh, and sigmoid. Out of a total of 50 terrains, 10 (20%) randomly selected terrains were used as the test set, and data from the remaining 40 (80%) terrains were used as the training set. The numbers of the total training set and test set were 280 and 70, respectively. A validation set was not separately divided since the BO method was applied to determine optimal architecture to prevent overfitting.

As indicators of prediction performance, coefficient of determination (R^2), root-mean-square error (RMSE), and maximum absolute error (MAE) were applied. R^2 is commonly used to assess the goodness of fit of surrogate models. If the model perfectly predicts the variance of the data, R^2 equals 1. This metric provides an overall indication of how well the model fits the data but can sometimes fail to reflect overfitting. RMSE directly measures the accuracy of predictions, and lower RMSE values indicate higher accuracy. It is sensitive to outliers, providing a measure of the magnitude of large errors. The MAE metric is also sensitive to outliers. Additionally, this can be beneficial to directly evaluate the worst case.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \tag{4}$$

$$MAE = \max(|O - P|) \tag{5}$$

3. Results

3.1. Non-Linear Fitting Model

If general non-linear regression shows better prediction performance, using the ANN method as a surrogate model might not be necessary. As a prelimanary anlaysis for computational effectiveness, the non-linear fit (NLF) using a 2nd order-polynomial function was conduted with the same parameters as PANN, i.e., $z_{0,eff}$, COV_{z_0}, and α. Similar to the ANN models, 80% of the dataset was used for NLF and the other 20% of the data was used as the test set.

The prediction performance of NLF models with varying patch window sizes is depected in Fig. 6, showcasing (i) R^2, (j) RMSE, and (k) MAE metrics. Fig. 6 (a) and (b) illustrate the performance for peak $C_{p,mean}$ and max $C_{p,RMS}$, respectively. The further to the bottom right of the heatmap, the larger the area that was considered when calculating $z_{0,eff}$ and COV_{z_0}. For example, the cell at the most bottom right indicates that 18 blocks in *y*-direction and 62 blocks in *x*-direction were used to calculate $z_{0,eff}$ and COV_{z_0}. Fig. 6 (a) shows the prediction performance for $C_{p,mean}$. On the other hand, fewer blocks are considered toward the upper left. The number of blocks considered changes around the location closest to the building model, that is, the block corresponding to (x, y)=(29500 mm, 0 mm) in Fig. 1. For example, the model of W×L=1×2 utilizes the two blocks located in the center of the row closest to the building model.

The overall prediction performance was acceptable in terms of the three performance indicators. The worst R^2 was still acceptable as 0.95 when W×L=18×62. It was clearly shown that the prediction performance improved with a smaller window size. When the patch size was within W×L=8×3, R^2 reached 0.99, and RMSE and MAE were also reduced to less than 0.04 and 0.16, showing excellent accuracy.

As shown in Fig. 6 (b), the NLF model for max $C_{p,RMS}$ also displayed higher prediction performance when the patch size was within W×L=8×3. However, its accuracy was relatively worse than the model for peak $C_{p,mean}$, showing an R^2 of less than 0.8. Additionally, prediction

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

performance cannot be guaranteed on roofs or leeward walls where flow separation occurs and greater wind pressure variability is observed. Conclusively, there were significant limitations in achieving acceptable prediction accuracy using NLF for all statistics of C_p on all walls. However, based on the change in prediction performance with varying patch size, it can be inferred that the morphology information within W×L=8×3 had the highest correlation with the wind pressure.

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

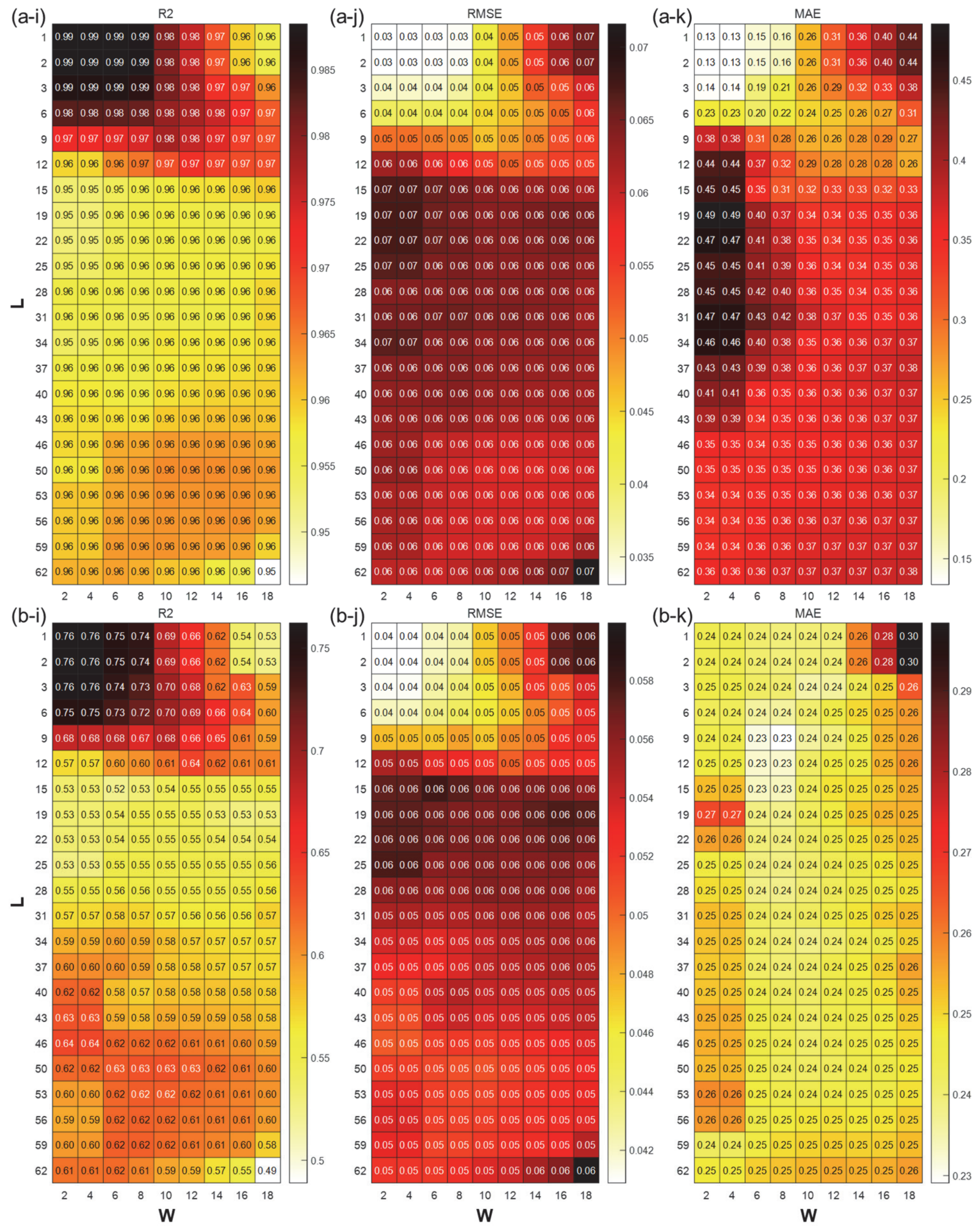


Fig. 6. Prediction performance of NLF in windward wall: (a) $C_{p,mean}$, and (b) $C_{p,RMS}$; with (i) R^2, (j) RMSE, and (k) MAE.

3.2. PANN and MANN Models

Figs. 7 and 8 showcase the prediction performances of PANN and MANN models. The ANN models were not developed for all patch sizes, and the considered patch sizes were limited within W×L=12×12. PANN demonstrated exceptional predictive accuracy for peak $C_{p,mean}$, achieving an R^2 nearing 1 and an MAE lowered to 0.07.

MANN also showcased impressive prediction performance. For peak $C_{p,mean}$, R^2 was improved to nearly 1.00, with RMSE and MAE reduced to 0.04 and 0.09, respectively. The predictions for max $C_{p,RMS}$ also indicated strong performance, as R^2 reached 0.92. However, the RMSE and MAE for MANN were slightly higher than those for PANN, highlighting that the empirical parameters in PANN capture terrain complexity more effectively. Moreover, with MANN, the increase in the number of input features with patch size could diminish computational efficiency, and its performance, in terms of MAE and RMSE, seems less optimal than that of PANN.

For both models, the predictions for max $C_{p,RMS}$ revealed a challenge, with lower R^2 and higher RMSE and MAE than for peak $C_{p,mean}$. This discrepancy underscores the complexities in accurately estimating wind pressure variability, leading to greater data dispersion.

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

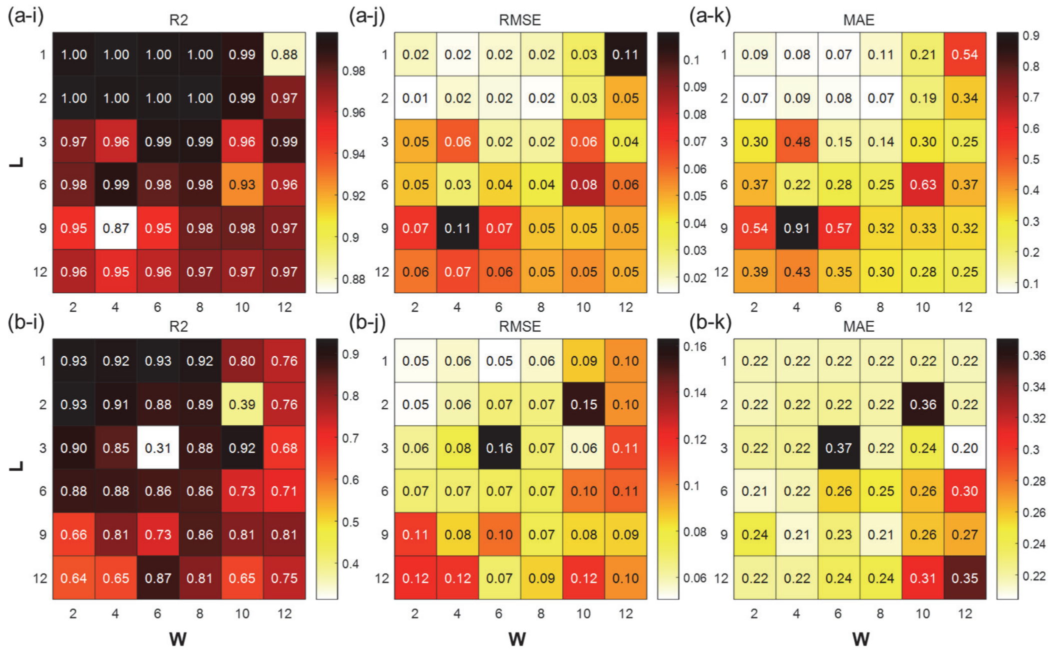


Fig. 7. Prediction performance of PANN models in windward wall: (a) Peak $C_{p,mean}$, and (b) Max $C_{p,RMS}$; with (i) R^2, (j) RMSE, and (k) MAE.

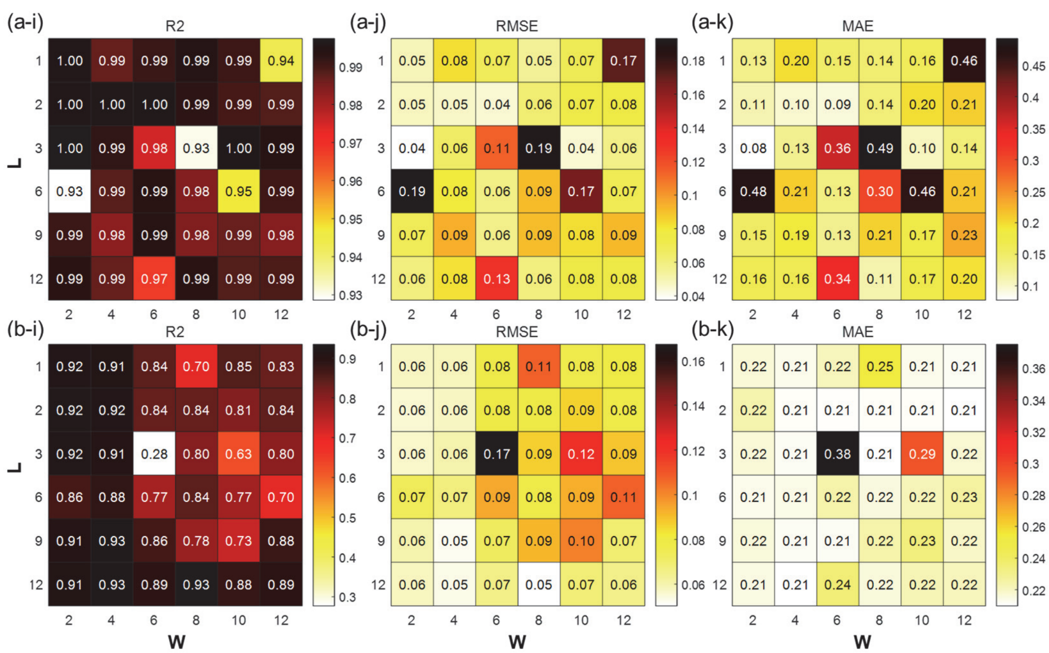


Fig. 8. Prediction performance of MANN models in windward wall: (a) Peak $C_{p,mean}$, and (b) Max $C_{p,RMS}$; with (i) R^2, (j) RMSE, and (k) MAE.

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

Fig. 9 displays the training histories for PANN and MANN models with a patch size of W×L=4×2, which demonstrated the best overall prediction performance. Since no validation set was used, only the loss for the training set is presented. The backpropagation algorithm was utilized to minimize the loss (MSE), and by the end of the training process, both PANN and MANN models converged to satisfactory performance results.

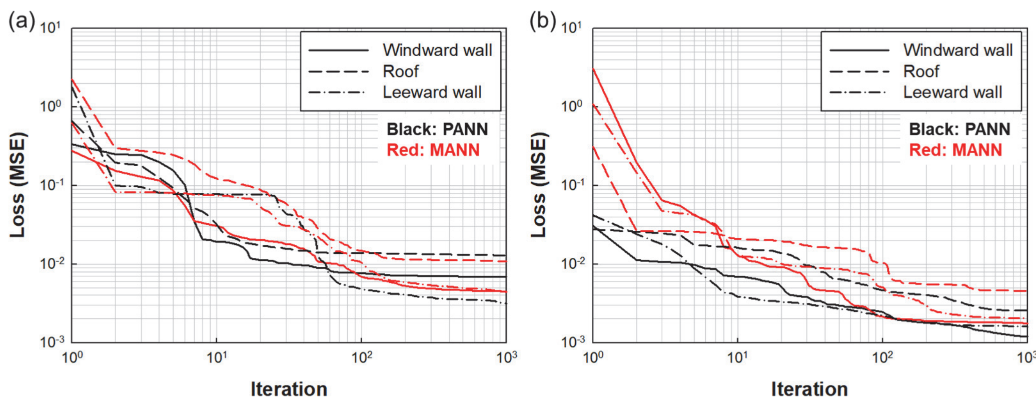


Fig. 9. Performance history of training sets: (a) Peak $C_{p,mean}$; and (b) Max $C_{p,RMS}$.

3.3. Comparison

Tables 1 and 2 detail the highest prediction accuracies for peak $C_{p,mean}$ and max $C_{p,RMS}$ for NLF, PANN, and MANN, along with the optimal patch sizes. NLF demonstrates robust prediction accuracy for $C_{p,mean}$ on the windward wall, attributed to its lesser susceptibility to variability-inducing phenomena like flow separation and vortices. This implies a lesser degree of nonlinearity, enabling NLF to achieve strong predictive outcomes. However, across the other five models—peak $C_{p,mean}$ for the roof and leeward wall, and max $C_{p,RMS}$ for the windward wall, roof, and leeward wall— the ANN models outperformed NLF in prediction accuracy.

Both PANN and MANN exhibited outstanding predictive accuracy with R^2 exceeding 0.9 in nearly all scenarios, barring max $C_{p,RMS}$ predictions for the leeward wall. The reduced performance for max $C_{p,RMS}$ on the leeward wall is attributed to significant wind pressure variability in this region, likely due to vortices. This variability presented challenges in achieving comparable prediction accuracy to that of the windward wall and roof, using only a limited set of empirical parameters ($z_{0,eff}$ and COV_{z_0}) or solely terrain morphology data.

Analysis of the optimal patch size for the best-performing model reveals that when terrain morphology information within a W×L=4×2 area was utilized, all three models−NLF, PANN, and MANN−showed highly satisfactory predictive results. Differences in performance among these models within this specific range were negligible. The simulated full-scale terrain area by these blocks, W×L=4×2, approximates 72 m × 23 m. Therefore, predictions of peak $C_{p,mean}$ and max $C_{p,RMS}$ with sufficiently high accuracy are possible if morphology information corresponding to at least a 100 m × 50 m area is acquired for actual complex heterogeneous terrains. This is approximately equivalent to 25 and 12.5 times the building height H (= 4 m). This suggests that the morphology within approximately $25H \times 12.5H$ in front of the WERFL low-rise building has the greatest correlation with the wind pressure coefficient, while terrain morphology at locations farther away from the low-rise building has a lower correlation with the wind characteristics and pressure experienced by the low-rise building. This is consistent with the results observed in previous studies [2, 5]. Consequently, when training an ANN model using information that includes distant terrain morphology, the model's performance may deteriorate.

Fig. 10 demonstrates the test set prediction results for NLF, PANN, and MANN models. With the exception of peak $C_{p,mean}$ for the windward wall, NLF frequently surpassed the 10% error margin, even revealing data points exceeding the 25% error bound. Conversely, both PANN and MANN maintained acceptable prediction accuracies.

Fig. 11 displays prediction results using NLF, PANN, and MANN at three specific sites−5, 31, and 43−randomly chosen from the test set. The morphologies of these three sites are shown in Fig.

12. Although the selection process for the test set was random, the chosen terrains successfully exhibit a wide range of real-world terrain complexity. As shown in Fig. 11, variations in terrain can significantly influence the pressure coefficient. Specifically, at site 43, which is characterized by relatively high terrain complexity as shown in Fig. 12 (c), there was a noticeable increase in larger values of $C_{p,mean}$ and $C_{p,RMS}$ due to increased turbulence intensity. This phenomenon has been detailed by An and Jung [6]. PANN and MANN accurately captured the trends of peak $C_{p,mean}$ and max $C_{p,RMS}$ as the wind incident angle varied. The nonlinearity of peak $C_{p,mean}$ on the windward wall was notably less than in the other five cases, which enables NLF to exhibit high prediction performance. However, for scenarios with enhanced nonlinearity, such as max $C_{p,RMS}$ on the windward wall and both peak $C_{p,mean}$ and max $C_{p,RMS}$ on the roof and leeward wall, NLF's performance lagged behind the ANN models.

Table 1. Comparison of best prediction performance for peak $C_{p,mean}$.

Wall	**Model**	**Patch size**		**Prediction performance**		
		W	**L**	**R^2**	**RMSE**	**MAE**
Windward wall	NLF	4	2	0.988	0.033	0.134
	PANN	4	2	0.996	0.019	0.093
	MANN	2	2	0.996	0.048	0.111
Roof	NLF	4	2	0.941	0.081	0.348
	PANN	4	2	0.986	0.039	0.315
	MANN	2	2	0.992	0.069	0.161
Leeward wall	NLF	4	2	0.811	0.079	0.236
	PANN	4	2	0.986	0.021	0.116
	MANN	2	2	0.992	0.039	0.112

Table 2. Comparison of best prediction performance for max $C_{p,RMS}$.

Wall	**Model**	**Window size**		**Prediction performance**		
		W	**L**	**R^2**	**RMSE**	**MAE**
Windward wall	NLF	4	2	0.762	0.041	0.244
	PANN	2	1	0.934	0.022	0.223
	MANN	2	2	0.921	0.055	0.223
Roof	NLF	4	2	0.849	0.041	0.185
	PANN	2	1	0.970	0.018	0.105
	MANN	2	1	0.934	0.064	0.137
Leeward wall	NLF	4	2	0.763	0.034	0.258
	PANN	2	1	0.834	0.028	0.267
	MANN	2	1	0.815	0.070	0.269

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

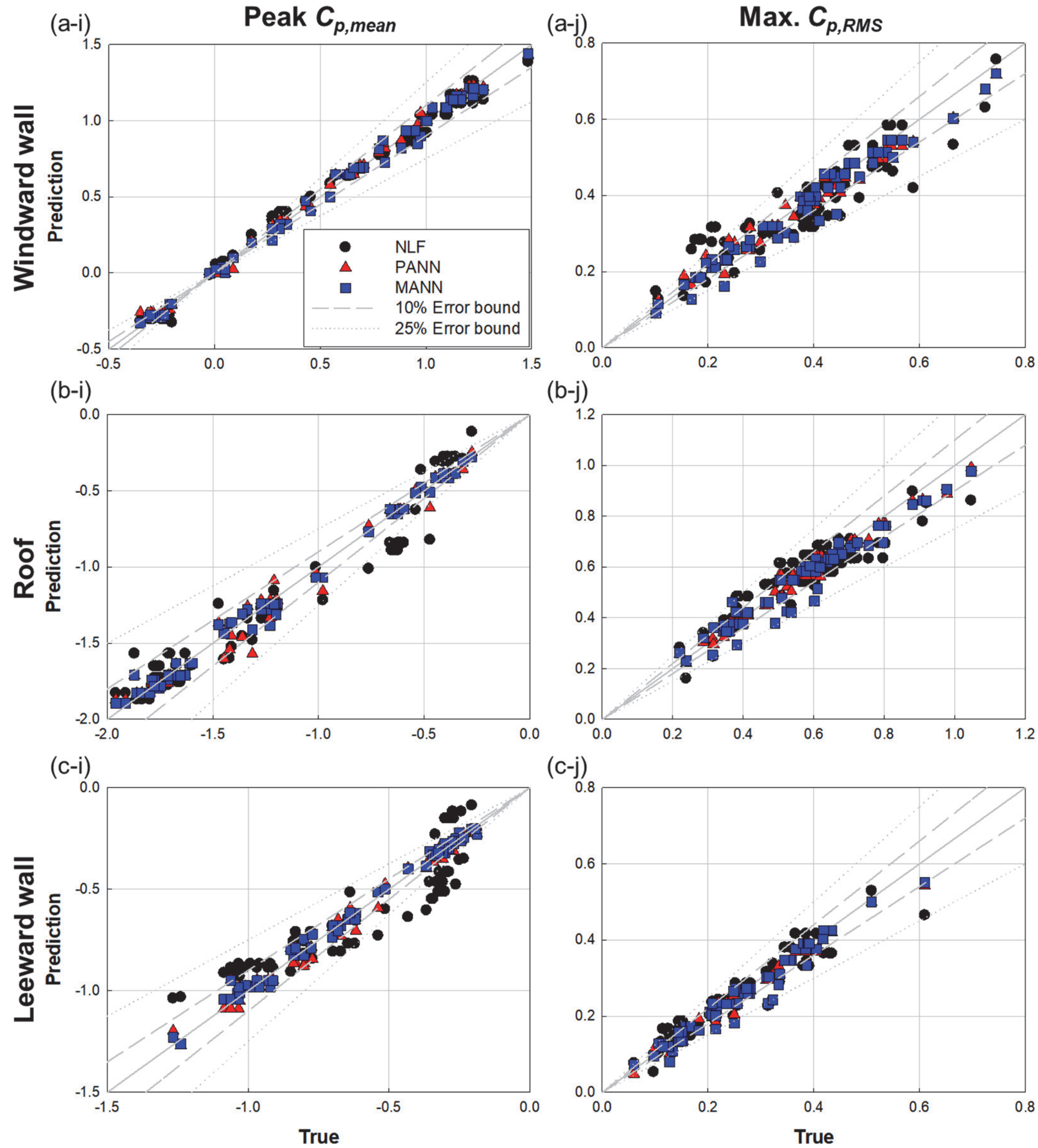
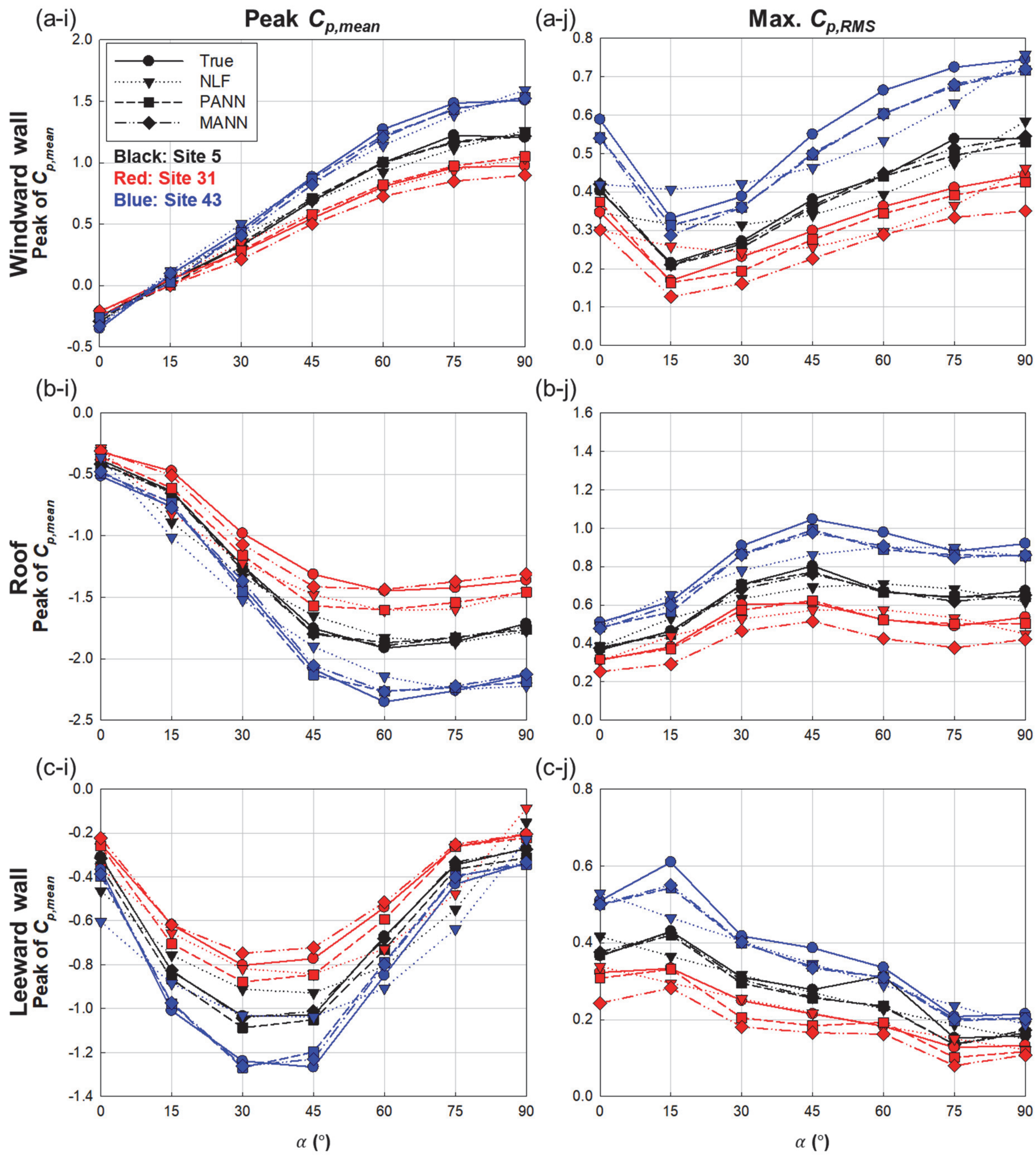


Fig. 10. Comparison of prediction results for NLF, PANN, and MANN for test set: (a) Windward wall, (b) Roof, and (c) Leeward wall; with (i) Peak $C_{p,mean}$, and (j) Max $C_{p,RMS}$.

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022



384

385 Fig. 11. Comparison of prediction results for specific sites in test set with varying wind incident angles: (a)
386 Windward wall, (b) Roof, and (c) Leeward wall; with (i) Peak $C_{p,mean}$, and (j) Max $C_{p,RMS}$.

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

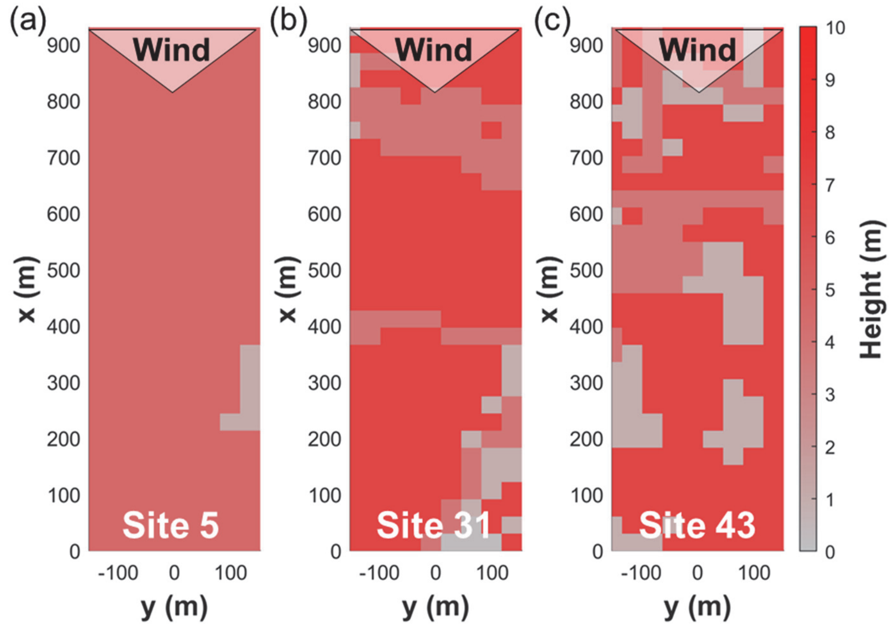


Fig. 12. Morphology of selected terrains used in test sets: (a) Site 5; (b) Site 31; and (c) Site 43.

Fig. 13 illustrates the differences between the predicted and actual values of peak $C_{p,mean}$ and max $C_{p,RMS}$ as a function of changes in $z_{0,eff}$. The $z_{0,eff}$ was calculated using the morphology information within a W×L=4×2 area. To avoid the potential misrepresentation of small difference values as disproportionately large errors when expressed as percentages, these difference values are presented. Across all $z_{0,eff}$ ranges, similar prediction accuracies are observed for both peak $C_{p,mean}$ and max $C_{p,RMS}$, with no significant differences identified within any specific $z_{0,eff}$ range. Histograms of the differences showcase that the majority of the distribution is tightly clustered around zero. $C_{p,RMS}$ values are typically smaller than $C_{p,mean}$ values. Consequentaly, it was observed that the scatter would be more concentrated within a narrower range of differences for $C_{p,RMS}$ compared to $C_{p,mean}$. This pattern is consistent with the MAE results presented in Tables 1 and 2.

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

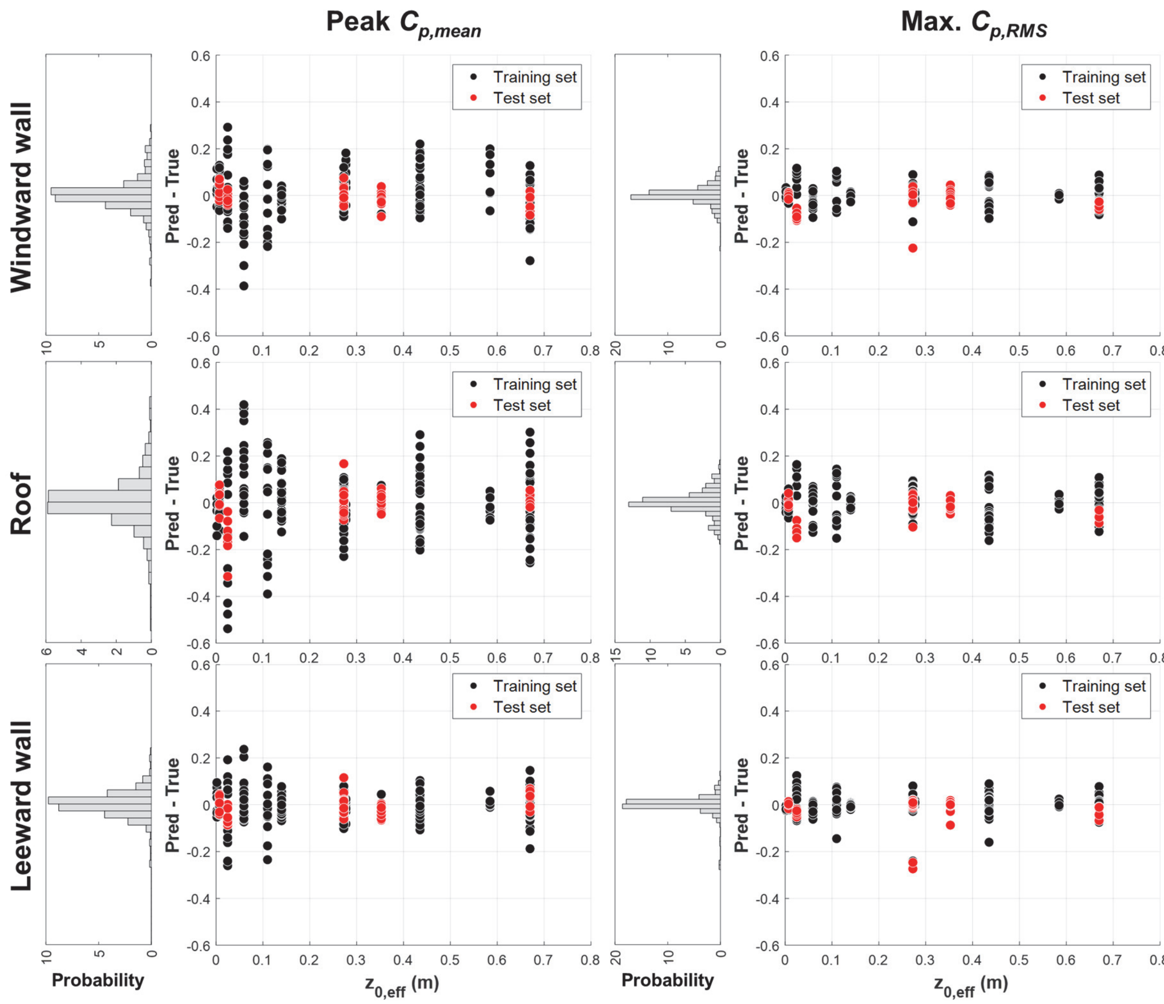


Fig. 13. The difference between predicted and true value of peak $C_{p,mean}$ and max $C_{p,RMS}$.

Sensitivity analysis evaluates how variations in a model's or system's inputs contribute to uncertainty in its outputs. Conducting sensitivity analysis on the developed ANN models allows us to identify the parameters that significantly influence wind pressure statistics. The Sobol method, a type of global sensitivity analysis, is particularly advantageous because it measures sensitivity across the entire input space, accommodates nonlinear responses, and assesses the effects of interactions in non-additive systems [43]. The Sobol index was calculated using the Saltelli method [44]. Fig. 14 outlines the global sensitivity analysis results. The PANN and MANN

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

models utilized terrain morphology information of W×L=4×2. The incident wind angle was omitted to concentrate on terrain complexity-related variables.

As illustrated in Fig. 14 (a), $z_{0,eff}$ emerged as a more influential factor than COV_{z_0} in all instances except for peak $C_{p,mean}$ at the Leeward wall, where, unlike other wall types, an increase in $z_{0,eff}$ did not correlated with heightened peak $C_{p,mean}$. This pattern, well-represented in the ANN models, aligns with observations by An and Jung [6]. Additionally, max $C_{P,RMS}$ was more significantly impacted by terrain roughness and complexity compared to peak $C_{p,mean}$, a finding consistent with the variable's direct association with wind speed variability. These sensitivity outcomes from the PANN model corroborate existing wind engineering insights, affirming the model's accurate reflection of physical phenomena.

For MANN, given the broader impact scope of row units over individual blocks, sensitivity analysis results for blocks 1 to 4 were averaged as row 1, and those for blocks 5 to 8 as row 2. Due to MANN's extensive input features, each input's influence was diminished, resulting in lower sensitivity compared to PANN. Predominantly, row 2's morphology, being closer to the building model, held more sway over the peak pressure coefficient statistics, notably tripling the impact on max $C_{p,RMS}$ for the windward wall and roof compared to row 1.

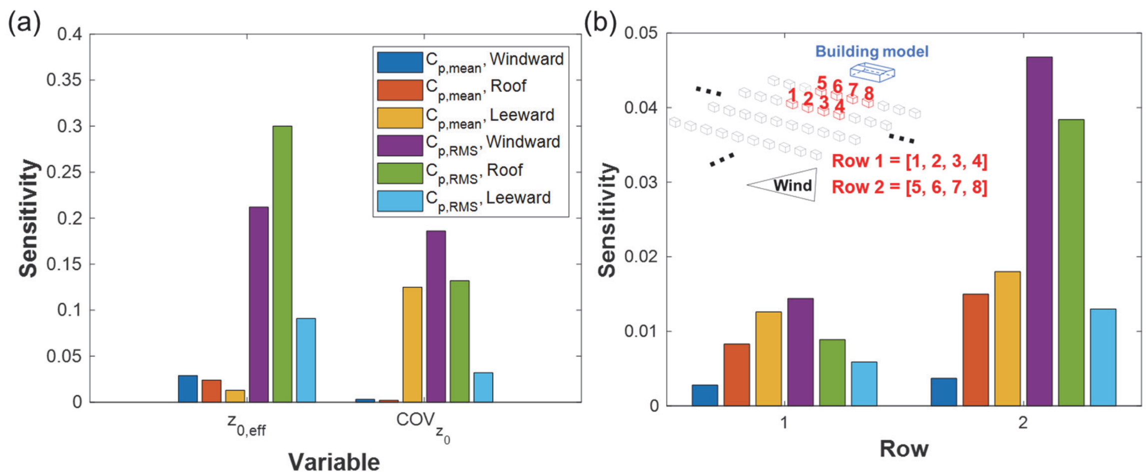


Fig. 14. Results of global sensitivity analysis for: (a) PANN, and (b) MANN.

4. Conclusions

This study introduced a data-driven approach to predicting peak values of pressure coefficient statistics for low-rise buildings situated in complex heterogeneous terrains. To address terrain complexity, we developed ANN models using two distinct sets of input features: empirical parameters-based ANN (PANN) and morphology-based ANN (MANN). We compared the prediction performance of these ANN models with that of a non-linear fitted (NLF) model and conducted global sensitivity analysis, yielding the following key insights:

- The NLF model demonstrated adequate prediction performance for peak $C_{p,mean}$ on the windward wall, attributed to the area's relatively lower non-linearity compared to other walls. However, NLF's efficacy diminished in other cases, such as max $C_{p,RMS}$ on the windward wall and both peak $C_{p,mean}$ and max $C_{p,RMS}$ on the roof and leeward walls, where PANN and MANN exhibited superior predictive accuracy.

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

- An examination of varying patch sizes revealed optimal prediction performance within a W×L=4×2 patch size, corresponding to a full-scale terrain area of approximately 72 m × 23 m. This means that the terrain morphology corresponding to about 100 m × 50 m (25*H* × 12.5*H*) has the strongest correlation with the wind pressure on the WERFL low-rise building. Therefore, securing terrain information exceeding 100 m × 50 m allows engineers to precisely predict wind pressure coefficients using the proposed ANN models.

- For max $C_{p,RMS}$ on the leeward wall, the R^2 values for NLF, PANN, and MANN were below 0.9, indicating reduced prediction performance for this area compared to the other five cases. The leeward wall experiences lower wind pressure and higher variability, often due to vortices, challenging the predictive accuracy of the models based on the input features utilized.

- Both PANN and MANN reflected the influence of wind incident angle variations and terrain complexity changes with reasonable accuracy across six outputs. The marginal differences in R^2, RMSE, and MAE between the two models suggest that the choice between PANN and MANN may depend on the available information during actual evaluations.

- Global sensitivity analysis underscored the greater impact of terrain roughness and complexity on max $C_{p,RMS}$ compared to peak $C_{p,mean}$ within the PANN model. Furthermore, $z_{0,eff}$ was identified as having a more significant influence than COV_{z_0} across all cases except for peak $C_{p,mean}$ at the leeward wall. In the MANN model, the block row nearest to the building model exerted a more pronounced effect on peak $C_{p,mean}$ and max $C_{p,RMS}$ than the subsequent row.

5. Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CMMI-1856205. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

6. References

[1] A.G. Davenport, Past, present and future of wind engineering, Journal of Wind Engineering and Industrial Aerodynamics 90(12-15) (2002) 1371-1380. https://doi.org/10.1016/S0167-6105(02)00383-5

[2] K. Wang, T. Stathopoulos, Exposure model for wind loading of buildings, Journal of Wind Engineering and Industrial Aerodynamics 95(9-11) (2007) 1511-1525. https://doi.org/10.1016/j.jweia.2007.02.016

[3] J. Yu, M. Li, T. Stathopoulos, Q. Zhou, X. Yu, Urban exposure upstream fetch and its influence on the formulation of wind load provisions, Building and Environment 203 (2021) 108072. https://doi.org/10.1016/j.buildenv.2021.108072

[4] Y.C. Kim, A. Yoshida, Y. Tamura, Characteristics of surface wind pressures on low-rise building located among large group of surrounding buildings, Engineering Structures 35 (2012) 18-28. https://doi.org/10.1016/j.engstruct.2011.10.024

[5] L. An, N. Alinejad, S. Kim, S. Jung, Experimental study on wind characteristics and prediction of mean wind profile over complex heterogeneous terrain, Building and Environment (2023) 110719. https://doi.org/10.1016/j.buildenv.2023.110719

[6] L.-S. An, S. Jung, Experimental investigation on influence of terrain complexity for wind pressure of low-rise building, Journal of Building Engineering (2024) 108350. https://doi.org/10.1016/j.jobe.2023.108350

[7] S. Kim, N. Alinejad, S. Jung, H.-K. Kim, The effect of open-to-suburban terrain transition on wind pressures on a low-rise building, Journal of Building Engineering (2024) 108651. https://doi.org/10.1016/j.jobe.2024.108651

[8] A.F. Akon, G.A. Kopp, Mean pressure distributions and reattachment lengths for roof-separation bubbles on low-rise buildings, Journal of Wind Engineering and Industrial Aerodynamics 155 (2016) 115-125. https://doi.org/10.1016/j.jweia.2016.05.008

[9] M. Kiya, K. Sasaki, Free-stream turbulence effects on a separation bubble, Journal of Wind Engineering and Industrial Aerodynamics 14(1-3) (1983) 375-386. https://doi.org/10.1016/0167-6105(83)90039-9

This file is the final accepted version of the manuscript, published in
https://doi.org/10.1016/j.buildenv.2024.112022

[10] X. Gavalda, J. Ferrer-Gener, G.A. Kopp, F. Giralt, Interpolation of pressure coefficients for low-rise buildings of different plan dimensions and roof slopes using artificial neural networks, Journal of wind engineering and industrial aerodynamics 99(5) (2011) 658-664.
[11] Y. Chen, G. Kopp, D. Surry, Interpolation of wind-induced pressure time series with an artificial neural network, Journal of Wind Engineering and Industrial Aerodynamics 90(6) (2002) 589-615. https://doi.org/10.1016/S0167-6105(02)00155-1
[12] F. Bre, J.M. Gimenez, V.D. Fachinotti, Prediction of wind pressure coefficients on building surfaces using artificial neural networks, Energy and Buildings 158 (2018) 1429-1441. https://doi.org/10.1016/j.enbuild.2017.11.045
[13] P.L. Fernández-Cabán, F.J. Masters, B.M. Phillips, Predicting roof pressures on a low-rise structure from freestream turbulence using artificial neural networks, Frontiers in Built Environment 4 (2018) 68. https://doi.org/10.3389/fbuil.2018.00068
[14] J. Tian, K.R. Gurley, M.T. Diaz, P.L. Fernandez-Caban, F.J. Masters, R. Fang, Low-rise gable roof buildings pressure prediction using deep neural networks, Journal of Wind Engineering and Industrial Aerodynamics 196 (2020) 104026. https://doi.org/10.1016/j.jweia.2019.104026
[15] Z. Ding, W. Zhang, D. Zhu, Neural-network based wind pressure prediction for low-rise buildings with genetic algorithm and Bayesian optimization, Eng. Struct. 260 (2022) 114203. https://doi.org/10.1016/j.engstruct.2022.114203
[16] L. Lang, L. Tiancai, A. Shan, T. Xiangyan, An improved random forest algorithm and its application to wind pressure prediction, International Journal of Intelligent Systems 36(8) (2021) 4016-4032. https://doi.org/10.1002/int.22448
[17] N. Alinejad, S. Kim, S. Jung, Wind-Tunnel Testing of Low-and Midrise Buildings under Heterogeneous Upwind Terrains, Journal of Structural Engineering 150(5) (2024) 04724001.
[18] N. Alinejad, S. Jung, G. Kakareko, P.L. Fernández-Cábán, Wind-Tunnel Reproduction of Nonuniform Terrains Using Local Roughness Zones, Bound. Layer Meteorol. (2023) 1-22. https://doi.org/10.1007/s10546-023-00822-0
[19] F.J. Masters, Boundary Layer Wind Tunnel, Basic Operations Manual, University of Florida, Gainesville, FL, 2017.
[20] R.A. Catarelli, P.L. Fernández-Cabán, B.M. Phillips, J.A. Bridge, F.J. Masters, K.R. Gurley, D.O. Prevatt, Automation and new capabilities in the university of Florida NHERI Boundary Layer Wind Tunnel, Frontiers in Built Environment 6 (2020) 558151. https://doi.org/10.3389/fbuil.2020.558151
[21] T. Stathopoulos, Wind loads on low-rise buildings: a review of the state of the art, Engineering Structures 6(2) (1984) 119-135.
[22] T.E. Ho, D. Surry, D. Morrish, G. Kopp, The UWO contribution to the NIST aerodynamic database for wind loads on low buildings: Part 1. Archiving format and basic aerodynamic data, J. Wind. Eng. Ind. Aerod. 93(1) (2005) 1-30.
[23] H.J. Ham, B. Bienkiewicz, Wind tunnel simulation of TTU flow and building roof pressure, Journal of Wind Engineering and Industrial Aerodynamics 77 (1998) 119-133.
[24] Y. Guo, C.-H. Wu, G.A. Kopp, A method to estimate peak pressures on low-rise building models based on quasi-steady theory and partial turbulence analysis, J. Wind. Eng. Ind. Aerod. 218 (2021) 104785.
[25] J. Wang, G.A. Kopp, Comparisons of aerodynamic data with the main wind force–resisting system provisions of ASCE 7-16. I: Low-rise buildings, Journal of Structural Engineering 147(3) (2021) 04020347. https://doi.org/10.1061/(ASCE)ST.1943-541X.0002925

[26] Y. Tamura, K. Suda, A. Sasaki, K. Miyashita, Y. Iwatani, T. Maruyama, K. Hibi, R. Ishibashi, Simultaneous wind measurements over two sites using Doppler sodars, Journal of Wind Engineering and Industrial Aerodynamics 89(14-15) (2001) 1647-1656. https://doi.org/10.1016/S0167-6105(01)00149-0
[27] A.N. .2, Australian/New Zealand standard, structural design actions. Part 2: wind actions, Standards Australia International Ltd.-Standards New Zealand, 2011.
[28] D. Surry, Pressure measurements on the Texas Tech building: wind tunnel measurements and comparisons with full scale, J. Wind. Eng. Ind. Aerod. 38(2-3) (1991) 235-247. https://doi.org/10.1016/0167-6105(91)90044-W
[29] Scanivalve, ZOC33 Miniature Pressure Scanner., 2023. http://scanivalve.com/products/pressure-measurement/miniature-analogpressure-scanners/zoc33-miniature-pressure-scanner/. (Accessed April 13rd 2023).
[30] M. Kovaerk, L. Amatucci, K.A. Gillis, F. Potra, J. Ratino, M.L. Levitan, D. Yeo, Calibration of dynamic pressure in a tubing system and optimized design of tube configuration: A numerical and experimental study, in: NIST (Ed.) Technical Note (NIST TN), National Institute of Standards and Technology, Gaithersburg, MD, 2018. https://doi.org/10.6028/NIST.TN.1994
[31] T. Ho, D. Surry, D. Morrish, NIST/TTU cooperative agreement–windstorm mitigation initiative: Wind tunnel experiments on generic low buildings, London, Canada: BLWTSS20–2003, Boundary-Layer Wind Tunnel Laboratory, Univ. of Western Ontario (2003).
[32] C. Homer, J. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. Herold, J. Wickham, K. Megown, Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information, Photogrammetric Engineering & Remote Sensing 81(5) (2015) 345-354. https://doi.org/10.14358/PERS.81.5.345
[33] D. Arthur, S. Vassilvitskii, K-means++ the advantages of careful seeding, Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, 2007, pp. 1027-1035.
[34] R. Macdonald, R. Griffiths, D. Hall, An improved method for the estimation of surface roughness of obstacle arrays, Atmospheric environment 32(11) (1998) 1857-1864. https://doi.org/10.1016/S1352-2310(97)00403-2
[35] N. Alinejad, S. Jung, G. Kakareko, P.L. Fernández-Cábán, Wind-tunnel reproduction of nonuniform terrains using local roughness zones, Bound. Layer Meteorol. 188(3) (2023) 463-484.
[36] T. Vihma, H. Savijärvi, On the effective roughness length for heterogeneous terrain, Quarterly Journal of the Royal Meteorological Society 117(498) (1991) 399-407. https://doi.org/10.1002/qj.49711749808
[37] P.A. Taylor, Comments and further analysis on effective roughness lengths for use in numerical three-dimensional models, Boundary-layer meteorology 39(4) (1987) 403-418. https://doi.org/10.1007/BF00125144
[38] R. Catarelli, P. Fernández-Cabán, F. Masters, J. Bridge, K. Gurley, C. Matyas, Automated terrain generation for precise atmospheric boundary layer simulation in the wind tunnel, Journal of Wind Engineering and Industrial Aerodynamics 207 (2020) 104276. https://doi.org/10.1016/j.jweia.2020.104276
[39] S. Jung, J. Ghaboussi, S.-D. Kwon, Estimation of aeroelastic parameters of bridge decks using neural networks, Journal of engineering mechanics 130(11) (2004) 1356-1364.
[40] R.N. D'souza, P.-Y. Huang, F.-C. Yeh, Structural analysis and optimization of convolutional neural networks with a small sample size, Scientific reports 10(1) (2020) 834.

[41] J. Snoek, H. Larochelle, R.P. Adams, Practical bayesian optimization of machine learning algorithms, Advances in neural information processing systems 25 (2012).
[42] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, N. De Freitas, Taking the human out of the loop: A review of Bayesian optimization, Proceedings of the IEEE 104(1) (2015) 148-175. 10.1109/JPROC.2015.2494218
[43] F. Cannavó, Sensitivity analysis for volcanic source modeling quality assessment and model selection, Computers & geosciences 44 (2012) 52-59. https://doi.org/10.1016/j.cageo.2012.03.008
[44] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Computer physics communications 181(2) (2010) 259-270.