Wind Engineering Joint Usage/Research Center
FY2025 Research Result Report

Research Field: Wind Hazard Mitigation
Research Year: FY2025
Research Number: 252004
Research Theme: Aerodynamic Interference Effects of Wind on Adjacent High-Rise Buildings: Machine Learning Models based on Wind Tunnel Data

Representative Researcher: Maciej Dutkiewicz

Budget [FY2025]: 360,000 Yen

1. Research Aim

The primary objective of this research is to evaluate the capability of numerical models to accurately predict Root Mean Square pressure coefficients (RMS Cp) with a focus on verifying the prediction accuracy of extreme values (max/min Cp) to ensure a reliable estimation of peak structural loads. The study introduces a novel weighted adaptive loss optimization approach to significantly improve the convergence and precision of pressure coefficient predictions. The research incorporates physics-informed analysis for Cp spatial reconstruction, enabling a more robust and physically consistent mapping of pressure across the entire surface.

2. Research Method

The study utilizes wind pressure data derived from the TPU database, focusing on a principal high-rise structure (70 m × 70 m × 280 m, scaled 1:400) subjected to aerodynamic interference from a neighboring building. A comparative evaluation of Machine Learning (ML) algorithms, including XGBoost, Deep Neural Networks (DNN), and Convolutional Neural Networks (CNN), is conducted to identify the most effective predictive architecture.

The Extreme Gradient Boosting (XGBoost) model utilizes decision trees as base learners to map complex non-linear aerodynamic interactions. Extreme Gradient Boosting is a powerful and efficient machine learning algorithm based on the gradient boosting framework. Boosting algorithms build multiple weak parts step by step, with each subsequent model learning from the errors of the previous ones. These models are then combined through weighted voting or summation, to produce a final, robust prediction.

LSBoost method implements least-squares boosting, where each decision tree minimizes the squared error (residuals) of the previous trees' predictions. The chart shown in Fig 1 illustrates the Boosting (Ensemble Learning) process used in the Spatially Regularized Gradient Boosting (SR-GB) model to reconstruct pressure fields. The sequence begins with the Original Data, which represents the measured pressure coefficient data obtained from wind tunnel tests for specific angles. Model 1 (Weak Learner), the first of decision trees, generates an initial, broad estimation of the pressure distribution. While Correctly Predicted Data refers to the portion where the model performed accurately, Weighted Errors identify the gaps where the first model failed. Within the algorithm, Laplacian smoothing-based spatial regularization is applied to ensure these errors remain physically consistent with the building's geometry.

The Weighted Data for Next Model step instructs the subsequent tree to focus specifically on rectifying the mistakes made by its predecessor. This iterative cycle continues through Model 2, Model 3, up to Model n for a total iterations, with each subsequent tree adding a small, calculated correction to the aggregate result. The Ensemble Model (Strong Learner) represents the final cumulative output of all trees. Although individual trees are considered weak, their weighted combination creates a powerful and highly accurate predictor. The Final Prediction constitutes the ultimate output: a comprehensive and physically smooth pressure profile.

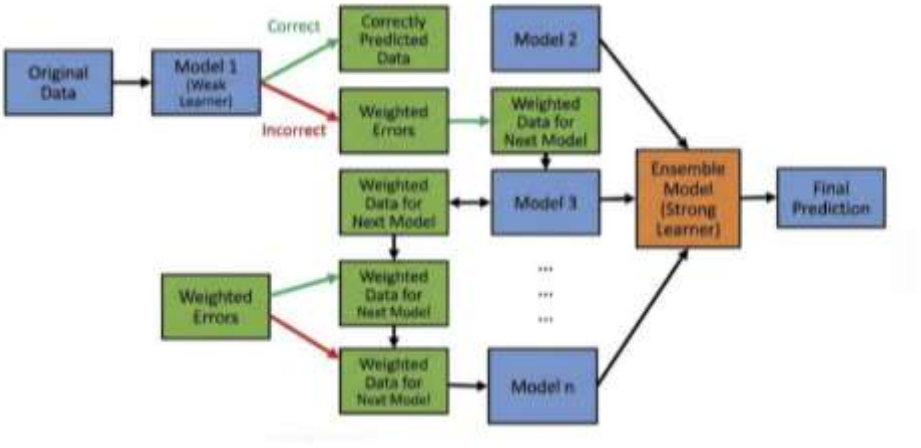
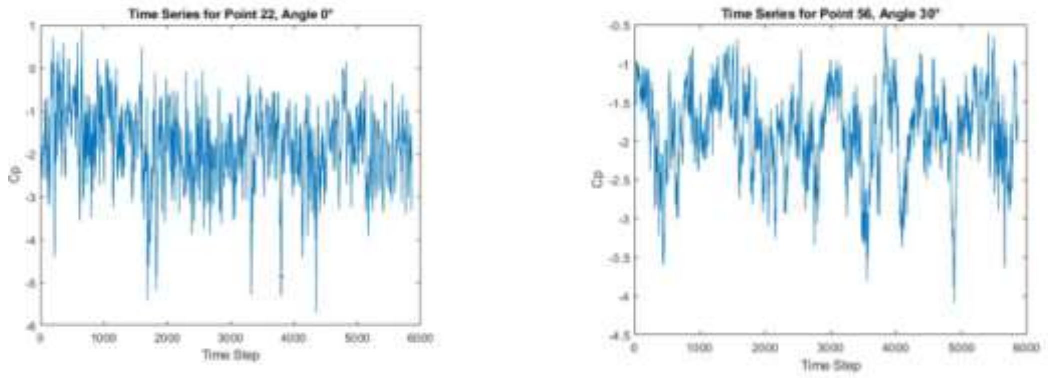


Fig.1 Boosting algorithm flowchart

Deep Neural Networks (DNN) represent a advanced class of connectionist systems that utilize multiple hidden layers to extract high-level features from complex, non-linear aerodynamic datasets . In wind engineering, DNNs are particularly effective at mapping the high-dimensional relationships between building configurations, wind incidence angles, and the resulting surface pressure coefficients recorded in the TPU database. The core of the Convolutional Neural Networks (CNN) architecture consists of convolutional layers that apply a set of learnable filters (kernels) to the input tensor. Each filter slides across the spatial dimensions (9×28) to compute feature maps. Novel DNN weighted adaptive loss optimization introduces the fundamental components of the proposed optimization framework. It reflects a strategic shift from standard machine learning practices toward a physically-consistent approach, specifically designed to achieve high pressure field fidelity. This methodology represents the original contribution of this research, distinguishing this method from standard Deep Learning approaches that typically fail to capture aerodynamic extremes. This methodology is fundamentally weighted, referring to the strategic assignment of varying importance to different data points; unlike standard models where every error is treated equally, this approach weights errors in high-pressure zones more heavily. Furthermore, the process is adaptive, indicating that the penalty mechanism is not static but dynamically adjusts its strength based on the actual magnitude of the pressure coefficient at each sensor location. Central to this architecture is the loss function, which defines the mathematical measure of discrepancy between the prediction and the experimental truth, making its minimization the primary objective of the training process. Finally, optimization describes the iterative process, which refines the network's internal parameters theta to find the optimal configuration where the model perfectly mirrors the complex aerodynamic behavior of the building.

By combining these elements, the algorithm ensures that the reconstructed pressure field maintains maximum fidelity, particularly in the leading-edge regions where leading-edge vortices generate the most critical structural loads. To overcome the smoothing limitations observed in standard Deep Neural Network architectures, an authorial modification to the backpropagation objective function is introduced. Standard regression layers often fail to capture extreme peaks, due to the global minimization of Mean Squared Error. The proposed algorithm utilizes a custom weighted MSE layer that dynamically scales the loss penalty based on the magnitude of the target pressure coefficient. By assigning higher weights to extreme experimental values, the optimizer is forced to prioritize the reconstruction of localized turbulence. This approach significantly reduces the spectral bias.

To evaluate the fundamental characteristics of the dataset used for machine learning training, the temporal pressure coefficient series were analyzed for critical sensor locations. This analysis focus on the experimental actual state to identify the fluctuations that the ML model must replicate (Fig.2).



(a) actual experimental pressure fluctuations for Sensor 22 at a wind angle of 0 degree.

(b) actual experimental pressure fluctuations for Sensor 56 at a wind angle of 30 degree.

Fig.2. Example of experimental pressure fluctuations for sensors: 22 and 56

3. Research Result

In addition to the standard 80/20 splitting methodology, where 80% of the samples are dedicated to the learning phase to recognize patterns and 20% are reserved for independent verification of accuracy and generalizability, the 10-fold cross-validation is used to evaluate the predictive model.

This approach minimizes the risk of overfitting and provides a more reliable estimate of the model's performance on unseen data. The term 10-fold R^2 cross-validation refers to an evaluation where the dataset is divided into 10 equal parts (folds), and the model is trained and tested 10 times, each time using a different fold as the test set and the remaining 9 folds as the training set.

This methodology ensures that every data sample is used exactly once for testing, while participating in the training process multiple times (Fig.3).

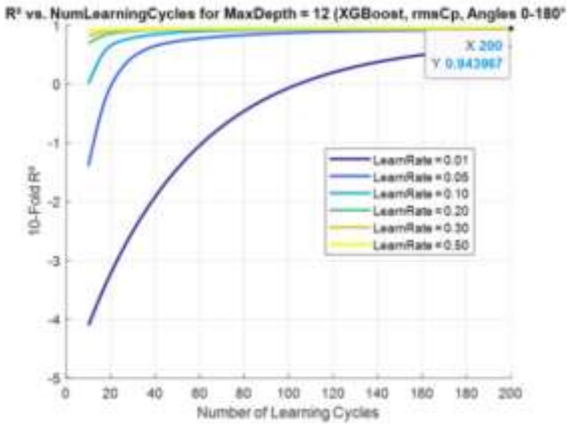
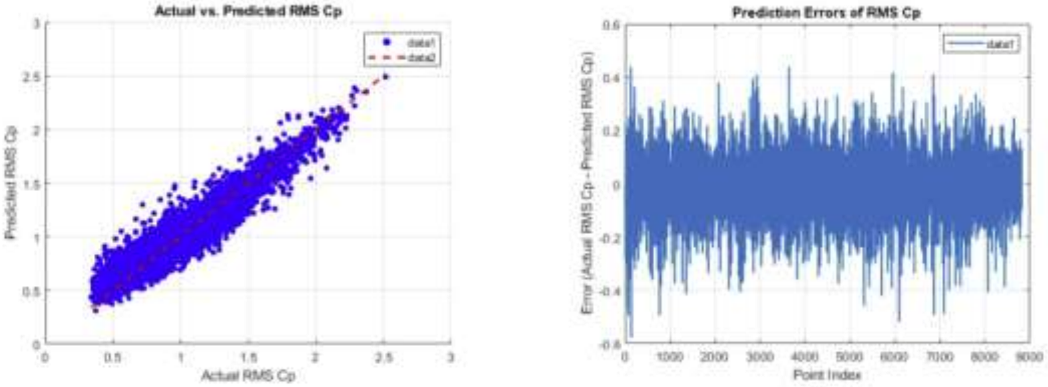


Fig.3. 10-fold R^2 performance across different learning rates and learning cycles.

The scatter plot illustrates the direct relationship between the actual experimental values and the predictions generated by the XGBoost model (Fig.4a). A strong concentration of data points along the 45-degree diagonal indicates a very high correlation and excellent fitting accuracy. Most data points are clustered near the ideal prediction line, which aligns with the 10-fold cross-validation R^2 result of 0.944. The low dispersion of points suggests that the model remains stable when predicting pressure fluctuations across different sensor locations. Such characteristics prove that the selected input features, including the vorticity proxy, correctly represent the underlying wind flow physics.



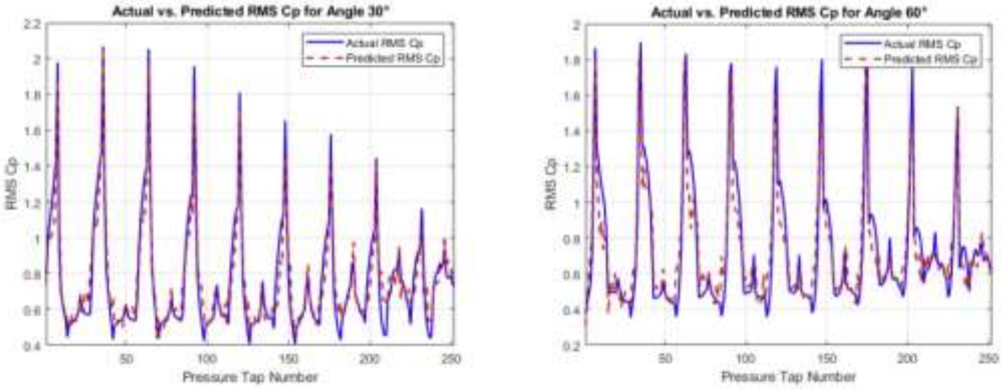
(a) scatter plot of actual versus predicted RMS Cp values for the XGBoost model.

(b) distribution of prediction errors (residuals) for RMS Cp results

Fig.4. Performance evaluation of the XGBoost model for RMS Cp prediction.

The prediction error plot (Fig.4b) enables a thorough assessment of the residuals and the identification of any potential systematic model biases. The error distribution is symmetric and clearly centered around zero, indicating an unbiased learning process. The narrow shape of the error curve confirms that the majority of predictions fall within a very small tolerance range relative to the experimental data. This analysis is crucial for verifying the stability of the LSBoost algorithm under conditions of high turbulence and aerodynamic interference. The limited occurrence of extreme errors demonstrates the high generalization capability of the XGBoost architecture on unseen datasets.

Figure 5 shows the comparison between actual and predicted RMS Cp values for wind angles of 30° and 60°.



(a) actual vs. predicted RMS Cp for wind angle 30°.

(b) actual vs. predicted RMS Cp for wind angle 60°.

Fig.5. Comparison of actual and predicted RMS Cp values for wind angles of (a) 30° and (b) 60°.

At higher wind angles, such as 30° the flow patterns become more complex due to increased turbulence and interference. Despite this complexity, the model maintains high accuracy. The RMS Cp distribution for the 60° wind angle demonstrates distinct periodic peaks across the sensor array. The XGBoost model accurately reproduces these sharp transitions, particularly at the leading edges where suction is most intense. Minor deviations are confined to the base regions of the pressure peaks where complex flow separation occurs. Overall, the high correlation confirms the model's ability to learn side-wall pressure distributions effectively.

4. Research Group

1. Representative Researcher
   Maciej Dutkiewicz, Bydgoszcz University of Science and Technology, Poland

2. Collaborate Researchers
1. Kim, Yong Chul, Tokyo Polytechnic University, Japan

6. Abstract (half page)

Research Theme:
Aerodynamic Interference Effects of Wind on Adjacent High-Rise Buildings: Machine Learning Models based on Wind Tunnel Data
Representative Researcher (Affiliation)
Maciej Dutkiewicz, Bydgoszcz University of Science and Technology, Poland

Summary

This research investigates the performance of advanced numerical models in predicting Root Mean Square pressure coefficients (RMS Cp) and extreme values (max/min Cp) for high-rise structures. Accurately estimating these parameters is critical for determining structural loads and ensuring building safety. Furthermore, the proposed methodology is particularly valuable in cases where sufficient experimental or laboratory data are unavailable to determine the wind loads. This is especially relevant when aerodynamic interference from neighboring structures causes these loads to deviate significantly from standard building code recommendations, rendering traditional normative approaches insufficient.

The study introduces a novel weighted adaptive loss optimization approach to significantly improve convergence and precision. Furthermore, the research incorporates physics-informed analysis for spatial reconstruction, enabling a more robust and physically consistent mapping of pressure across the entire building surface. The study utilizes wind pressure data derived from the TPU database, focusing on a principal high-rise structure (70m×70m×280 m, scaled 1:400) subjected to aerodynamic interference from a neighboring building. A comparative evaluation of Machine Learning (ML) algorithms, including XGBoost, Deep Neural Networks (DNN), and Convolutional Neural Networks (CNN), is conducted to identify the most effective predictive architecture. The models are trained on a limited set of experimental data to explore the potential of ML in interpolating complex pressure fields under varying wind angles and building configurations. The results demonstrate that the proposed ML models exhibit a superior ability to map complex spatial pressure distributions caused by aerodynamic interference—phenomena that traditional linear models often fail to capture. The weighted adaptive loss optimization effectively addressed the data imbalance inherent in wind engineering by significantly reducing prediction errors in high-gradient and peak pressure zones. By integrating physical constraints into the learning process, the models achieved high fidelity in extremal pressure analysis. The proposed framework allows for the efficient densification of sparse wind tunnel data, offering a cost-effective tool for design stages by reducing the need for extensive experimental campaigns. The hybrid approach—combining the efficiency of XGBoost with the depth of CNN and Weighted DNN—provides a versatile tool for analyzing interference effects in diverse urban configurations. This methodology serves as a critical foundation for further research into multi-building systems, where wake-induced turbulence and localized pressure magnifications significantly impact structural integrity and urban safety.