Journal of Research and Applications in Mechanical Engineering
ISSN: 2229-2152 (Print); 2697-424x (Online)
(2025) Vol. 13, No. 3, Paper No. JRAME 25-13-035
[DOI: 10.14456/ jrame.2025.35]

Research Article

Noise Prediction of Cylinder Flow using Machine Learning

N. Chinaso
K. Rattanamongkhonkun
W. Rojanaratanangkule*
Department of Mechanical engineering Faculty of Engineering, Chiang Mai University, Chiang Mai, 50200, Thailand

Received 15 September 2024
Revised 25 November 2024
Accepted 14 December 2024

Abstract:
The accurate prediction of aerodynamic noise generated by cylinder flow is a critical challenge in various engineering applications, including automotive and aerospace industries. Traditional Computational Fluid Dynamics (CFD) methods, such as Direct Numerical Simulation (DNS), often require significant computational resources and time to simulate the complex interactions within flow. This study successfully creates a model using Machine Learning (ML) techniques to predict the pressure fluctuation in flow over a cylinder which provides a faster and equally reliable alternative to conventional methods.

Keywords: *Acoustic noise, Cylinder flow, Machine learning, Compressible, Low Mach number flow*

1. Introduction

In various industries, such as the automotive and aerospace sectors, cylinders are commonly encountered shapes and are frequently used as benchmarks in research due to the availability of reliable experimental data. Predicting the acoustic noise generated by fluid flow around objects like cylindrical structures is crucial, as this noise can impact performance and safety. However, calculating and predicting this flow-induced noise using traditional CFD methods, such as DNS, are complex and require substantial computational resources and time. This has created a strong demand for new techniques that can predict acoustic noise more efficiently while maintaining accuracy [1].

While DNS offers high-fidelity insights, it faces critical limitations in computational efficiency, posing significant challenges for industrial applications. Since DNS requires highly refined meshing and extensive computational resources, it is impractical for large-scale applications or situations requiring real-time predictions [2]. A review of related work reveals that multiple techniques have been developed to address this, such as combining Large Eddy Simulation (LES) with acoustic models. However, these methods do not provide high enough accuracy in noise predictions [3].

In response to these limitations, ML has been explored as a promising alternative for predicting flow-induced noise. ML models are capable of bypassing the intricate calculations required in DNS, instead of learning from large datasets of simulations or experimental results to capture complex noise generation patterns. ML techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) can capture both spatial and sequential flow information, and studies indicate that ML can achieve results comparable to DNS while reducing computational cost by over 90% [4]. Moreover, ML models can be tailored to specific conditions, such as varying flow velocities, geometries, and boundary conditions, to improve prediction accuracy [5].

Recent studies, such as Mosavi et al., demonstrated that ML algorithms could reliably predict flow-induced noise around bluff bodies like cylinders, achieving up to a 90% reduction in computational cost [6]. Additionally, Wang et al. developed hybrid ML models incorporating physical constraints, that enhanced accuracy while reducing the risk of overfitting—a common issue in purely data-driven approaches [7].

* Corresponding author: W. Rojanaratanangkule
E-mail address: watchapon.roj@cmu.ac.th

This research aims to explore the potential of ML in creating models to predict acoustic noise generated by fluid flow around a cylinder, focusing on the accuracy of the models compared to the widely used DNS methods. Additionally, ML has the potential to significantly reduce computational costs by bypassing the intensive calculations required in traditional CFD approaches [8]. The results of this research will contribute to the development of highly efficient and cost-effective acoustic noise prediction tools.

2. Numerical Approach

2.1. Datasets

Data of flow past a cylinder was prepared and collected from a DNS conducted at a Reynold number of 150 and a Mach number of 0.2. These values were chosen to ensure a balance between computational feasibility and physical relevance to flow-induced noise prediction. At a Reynolds number of 150, the cylinder flow exhibits the laminar vortex shedding. This phenomenon is ideal for investigating flow-induced noise because it produces periodic pressure fluctuations that generate significant acoustic noise [9]. The Mach number of 0.2 corresponds to a subsonic flow regime, where compressibility effects are minimal but still relevant to the study of aeroacoustics. At this speed, the generated acoustic noise remains unaffected by shock waves or supersonic flow instabilities, providing a clearer assessment of the noise generated purely from pressure fluctuations and vortex shedding [10]. The number of grid points in radial and azimuthal directions (n_r, n_θ) are 711 and 513. The dataset consists of 150 time series snapshots over 30 seconds. In this research, the denoted input X is constructed from flow property of pressure (p) to perform model prediction output Y for the instantaneous fluctuating pressure (p') and the next time steps of p'. Specifically, the model is trained using 100 snapshots of the pressure data p. After training, the model's performance is evaluated by testing it on 50 snapshots of p'. In this study, the training process is conducted for each grid point sequentially, utilizing a loop-based iteration through all grid points. This approach ensures that the model can effectively generalize its predictions for unseen data, capturing the intricate and evolving dynamics of the fluctuating pressure, p'over time. Hence, it enables more accurate forecasting of pressure variations even in complex and previously unobserved flow conditions.

Table 1: Details of DNS

Re_∞	150
Ma_∞	0.2
Grid points of n_r	711
Grid points of n_θ	513
Number of snapshots	150

The computational domain for simulating aerodynamic noise is separated into two zones, the physical and buffer zones. In this work, the buffer zone will not be considered in the analysis, as the data collected from the buffer zone has minimal significance compared to the physical zone. The buffer zone's primary function is to filter out acoustic disturbances and minimize the reflection of waves back into the computational domain [10]. Within the physical zone, a very fine grid resolution was defined, with the smallest grid spacing of approximately $\Delta r = 0.005D$ in the radial direction. This grid spacing was kept equidistant for approximately 101 grid points near the wall boundary condition to accurately capture the pressure field. For the azimuthal direction, 513 grid points with uniform grid spacing were applied over zone, as illustrated in Figure 1.

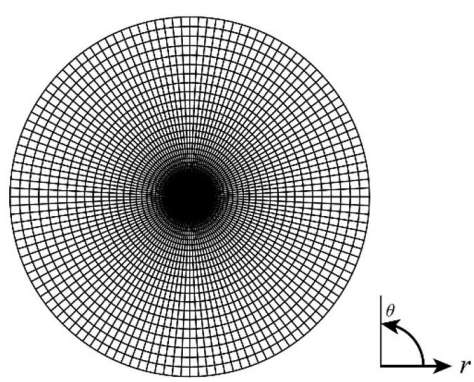


Fig. 1. Overall details of O-grid mesh used in this work.

The compressible fluid dynamics, three fundamental equations govern the flow: the conservation of mass, the conservation of momentum and the conservation of energy. These equations account for the variations in density, velocity, pressure, and temperature in compressible flows [11]. The conservation of mass for compressible flow is

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0 \tag{1}$$

where ρ is the fluid density, $\boldsymbol{u}$ is the velocity vector and t is time. The conservation of mass captures the variation of density at a specific location over time, which is fundamental in compressible flow dynamics [12].

The conservation of momentum for compressible flow, which describes the balance of forces, is given by

$$\frac{\partial (\rho \boldsymbol{u})}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}\boldsymbol{u}) = -\nabla p + \nabla \cdot \tau_{ij} + \rho \boldsymbol{f} \tag{2}$$

where p is the pressure, $\boldsymbol{\tau_{ij}}$ is the viscous stress tensor and $\boldsymbol{f}$ presents external body forces. This equation is vital for understanding how the fluid's momentum changes due to pressure, viscous forces, and external forces [13].
The energy equation accounts for the conservation of total energy (including internal, kinetic, and potential energies) in a compressible fluid flow is

$$\frac{\partial (\rho E)}{\partial t} + \nabla \cdot (\boldsymbol{u}(\rho E + p)) = \nabla \cdot (\boldsymbol{u} \cdot \tau_{ij}) + \nabla \cdot (k \nabla T) + \rho \boldsymbol{f} \cdot \boldsymbol{u} \tag{3}$$

where E is the total energy per unit mass, T is the temperature and k is the thermal conductivity. This equation captures the transport of energy due to conduction, viscous dissipation, and external work, which is particularly important in high-speed flows where compressibility effects dominate [12].

The compressible Navier–Stokes equations for conservative variables are solved in curvilinear coordinates via an in-house code, HiPSTAR. The flow solver employs a 4th-order optimized compact central finite-difference scheme [13] to discretize all the spatial derivatives in the radial (r) and azimuthal ($\emptyset$) directions. Temporal advancement is obtained by an ultra-low-storage five-step 4th-order Runge–Kutta scheme [14]. To enhance the stability of the code, a skew-symmetric splitting scheme of the convective terms and a pentadiagonal compact filter [15], applied after every full time-step with a cut-off wavenumber of 0.88, are used.

The relationship between the pressure $p(x, y, t)$ and the instantaneous fluctuating pressure $p'(x, y, t)$ is linear. This implies that p' can be expressed as a linear function of p, represented by the following equation

$$p'(x, y, t) = p(x, y, t) - \bar{p} \tag{4}$$

where $\bar{p}$ is the mean pressure is calculated as the average pressure p over the entire period of interest, and it serves as a baseline against which fluctuations are measured. This linear relationship is fundamental in the analysis of turbulent flows, where the fluctuating components of pressure are of primary interest, particularly in the study of flow-induced forces and noise generation. Following the prediction of p', the subsequent step involves the computation of its Root Mean Square (RMS) value, represented by p'_{rms}. This calculation is performed using the following equation

$$p'_{rms} = \sqrt{\overline{p'^2}} \tag{5}$$

where $\overline{p'^2}$ represents the time-averaged value of the square of the p' [16]. This term is calculated by taking the square of the p' at each time step, then averaging these squared values over the entire time period.

2.2. Linear regression

This study utilizes basic supervised learning techniques, specifically regression. Linear regression is a supervised learning algorithm that forms the basis for many advanced techniques in machine learning [17], and can be used to

predict a continuous dependent variable *Y* based on one or more independent variables *X*. The hypothesis function ($h_\theta(X)$) in linear regression is used to predict the output value *Y* for a given input *X*. It is defined as

$$h_\theta(X) = \theta_0 + \theta_1 X \tag{6}$$

where θ_0 *and* θ_1 is the vector of parameters, and *X* is the vector of input feature (including the intercept term).

In this formulation, the goal of linear regression is to find values of the parameters θ_0 and θ_1 that minimize the difference between the predicted values $h_\theta(X)$ and the actual observed value in the training data. This difference is typically measured using the cost function.

2.3. Cost Function

The cost function or loss function used in polynomial regression is the Mean Squared Error (MSE) function, which measures the average squared difference between the predicted values and the actual values. The cost function $J(\theta)$ is defined as

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=0}^{m-1}(h(X_i) - Y_i)^2 \tag{7}$$

where m is the number of training examples, X_i is the i-th input feature, and Y_i is the i-th output value [17].

2.4. Gradient Descent

Gradient descent is an optimization algorithm used to minimize the cost function by iteratively updating the parameters θ_j (where j = 0, 1). The update rule for gradient descent in polynomial regression is

$$\theta_j = \theta_j - \alpha\frac{\partial J(\theta_j)}{\partial \theta_j} \tag{8}$$

where α is the learning rate and $\frac{\partial J(\theta_j)}{\partial \theta_j}$ is the partial derivative of the cost function with respect to θ_j [17]. The gradient descent algorithm continues to adjust θ until the cost function converges to a minimum value [18, 19].

3. Results and Discussion

3.1. Results of training model

The performance of the trained model is evaluated through the cost function (*J*) plot over 800 epochs. The choice of 800 epochs is determined based on the observation that the cost function must reach a stable value prior to this point. As a result, further training beyond 800 epochs does not lead to significant improvements in model performance. This suggests that 800 epochs are sufficient to achieve convergence and adequately train the model, ensuring efficient use of computational resources without overfitting. The cost function, which quantifies the discrepancy between predicted and actual values, exhibits a pronounced decline throughout the training process. This decrease reflects effective optimization of the model parameters.

The model is trained with a learning rate of 1×10^{-2}, as this value ensures the convergence of the cost function across all iterations. Additionally, experiments have been conducted with lower learning rates, which yielded comparable the cost function values. However, these lower learning rates significantly increased the training time, making them less computationally efficient. Therefore, the learning rate of 1×10^{-2} is an optimal balance between convergence performance and computational cost. Initially, the cost function decreases rapidly, indicating substantial improvements in the model's accuracy. As training progresses, the rate of decrease in *J* is much slower, suggesting that the model approaches a convergence point with optimized parameters [17, 18]. The final cost function value is significantly lower than its initial value, demonstrating that the learning rate is appropriately set to balance the speed and stability of convergence, shows in Figure 2. This reduction in the cost function underscores the model's capability to learn from the data effectively and accurately predict p' from the input p.

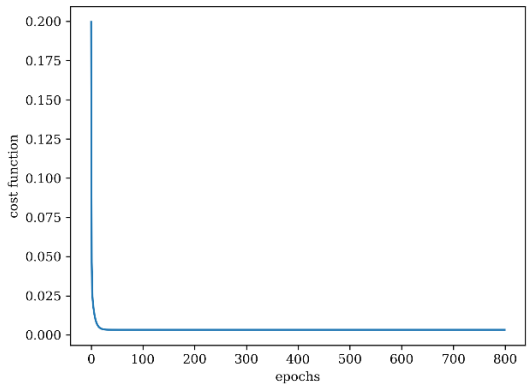


Fig. 2. The cost function (J) over the training epochs.

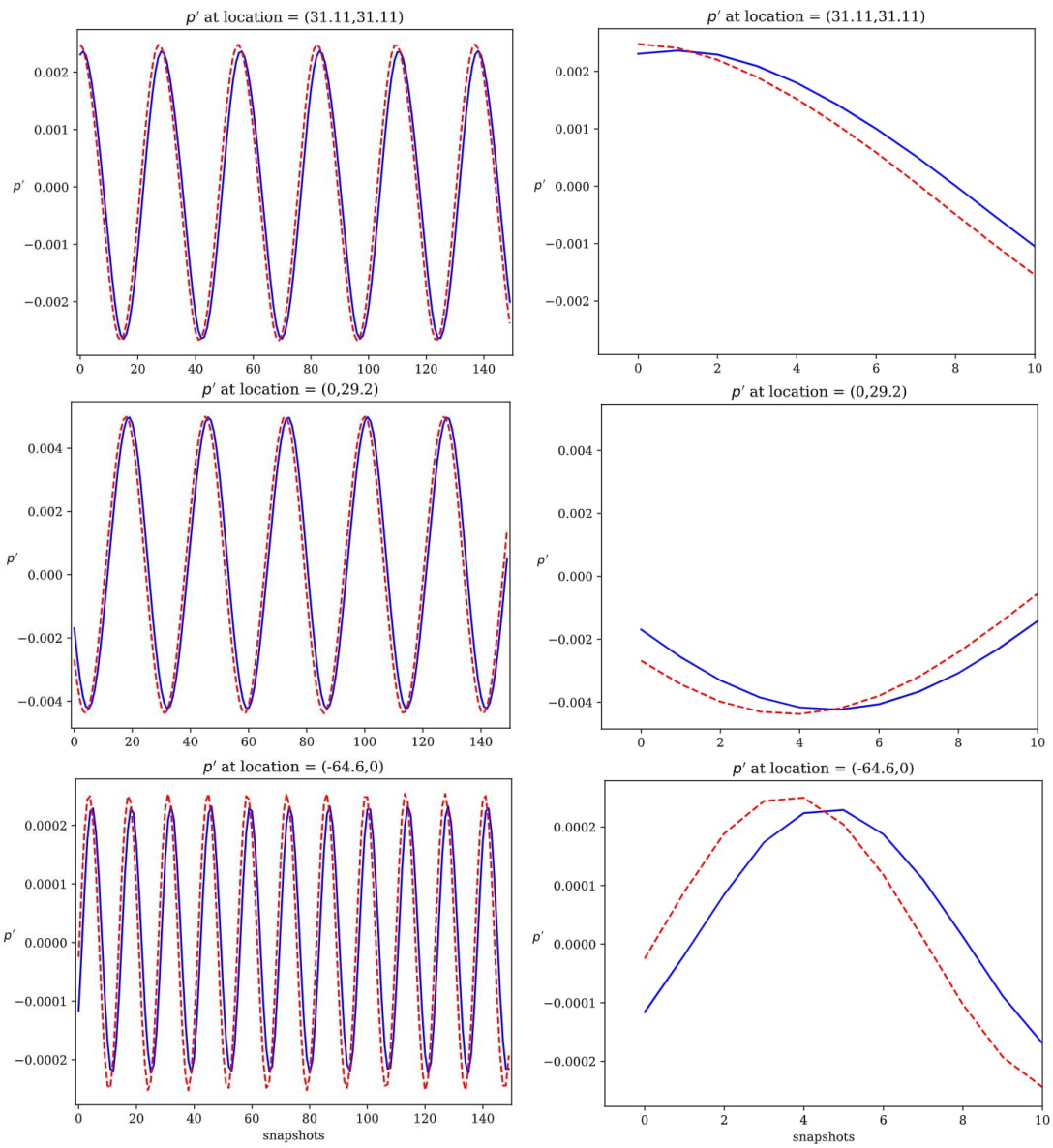


Fig. 3. The comparison between p'_{actual} (blue dots) and $p'_{predicted}$ (red dots) over a span of 150 snapshots at location of points.

The model is initially trained using a dataset comprising 100 snapshots with a total of 364,743 data points. This extensive training allows the model to learn and capture intricate patterns in p' effectively. Following the training phase, the model is evaluated on a separate test set consisting of 50 snapshots. The results of this evaluation are depicted in Figure 3, demonstrating a high level of agreement between the predicted and actual values. This indicates that the model is able to make accurate predictions of the unseen data.

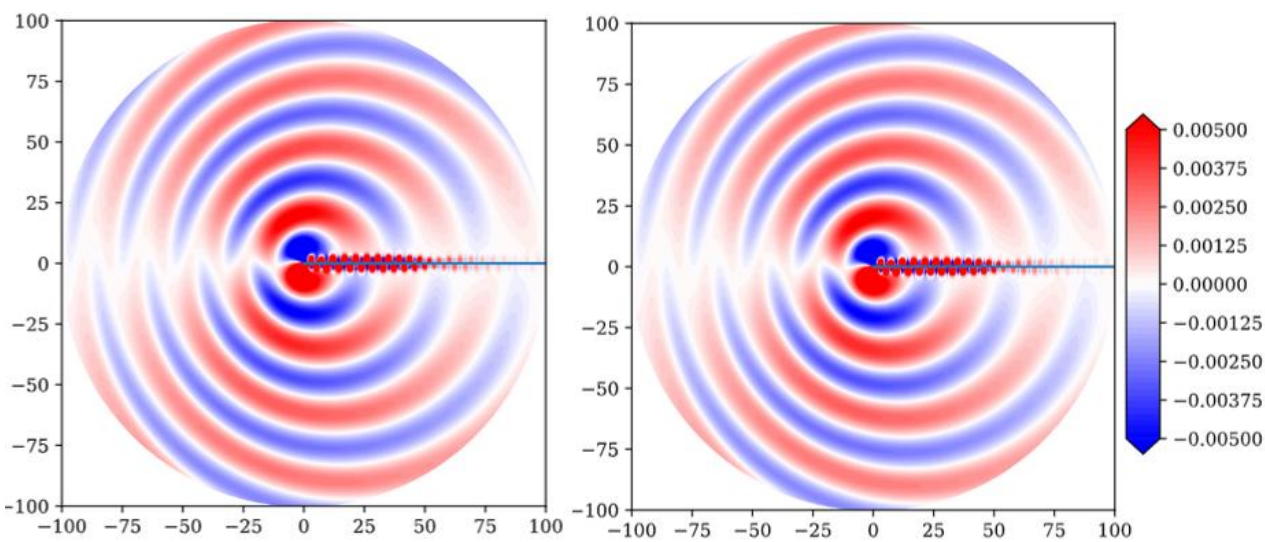
Specifically, the close alignment between p'_{actual} and $p'_{predicted}$ shows that the model is effectively captures the temporal dynamics of p'. The slight deviations are consistent with the inherent variability of the system and fall within acceptable error margins. This performance highlights the model's robustness and its ability to provide reliable forecasts of pressure fluctuations based on the training data. Overall, the results validate the model's capability to accurately predict p' over extended periods, underscoring its potential utility in practical applications.

3.2. The pressure fluctuation

The pressure fluctuation arises due to the fluid flow past a circular cylinder [1, 9], creating complex flow patterns that result in pressure variations on the cylinder's surface. Figure 4(a) illustrates a plot of p'_{actual}, providing a detailed view of the pressure behavior around the cylinder. In contrast, Figure 4(b) shows $p'_{predicted}$, plotted for comparison with the actual data presented in Figure 4(a).

In both contour plots, the color scheme is used to represent different pressure levels, the red color indicates positive pressure, while the blue color indicates negative pressure. The alternating colors between red and blue in these figures highlight regions of rapid pressure change, signifying the pressure fluctuations caused by the flow separation and vortex shedding around the cylinder. A visual inspection reveals that the two plots exhibit remarkable similarity, indicating that the model's predictions closely resemble the actual pressure fluctuations.

To further examine the differences between the actual and predicted p', the pressure values at a specific angular location of 90 degrees are extracted and analyzed, as shown in Figure 5 This analysis helps identify subtle discrepancies between the predicted and the actual pressure values, providing insight into the model's prediction accuracy.



(a) p'_{actual} plot. (b) $p'_{predicted}$ plot.

Fig. 4. Contour of the instantaneous pressure fluctuations.

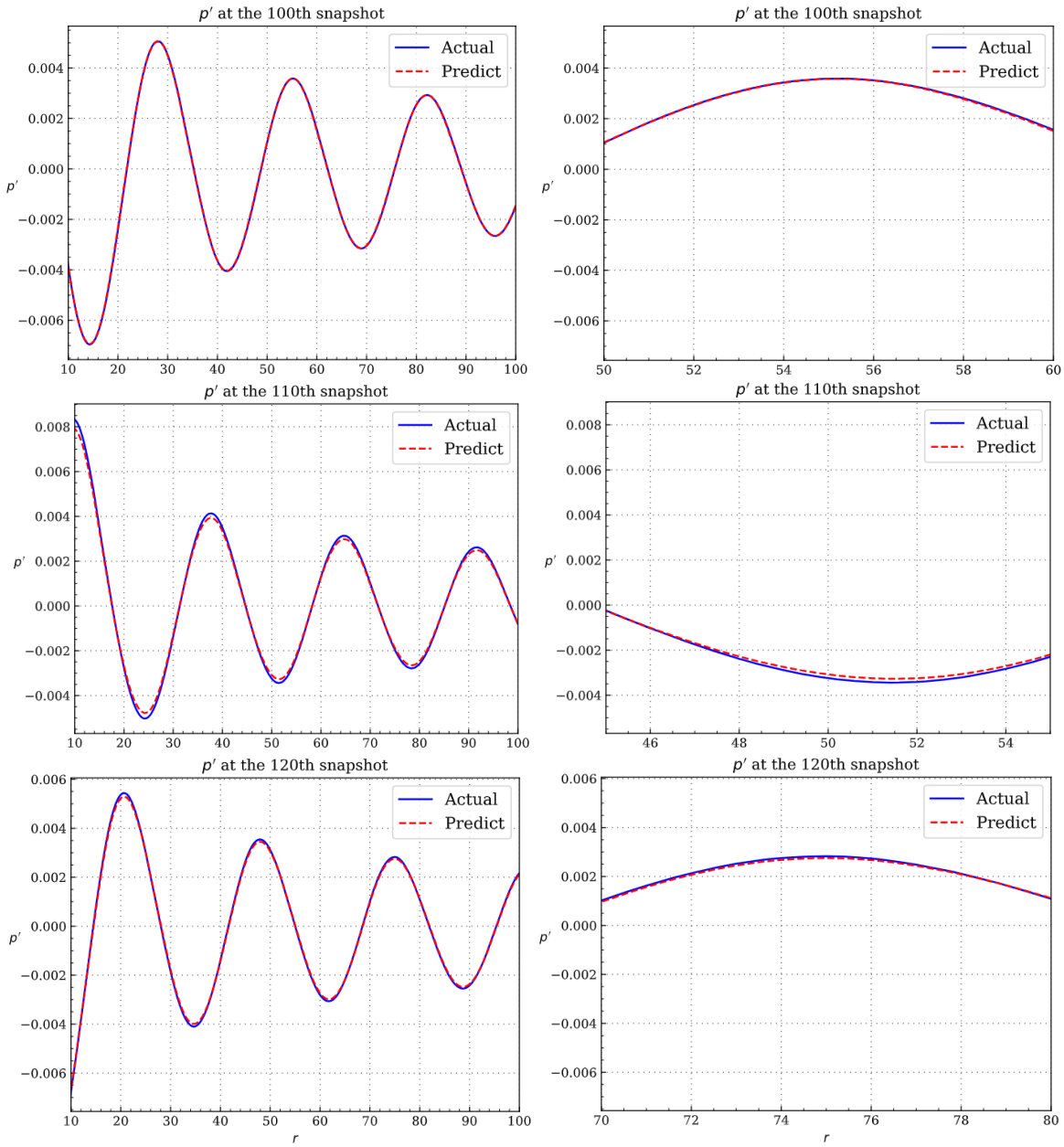
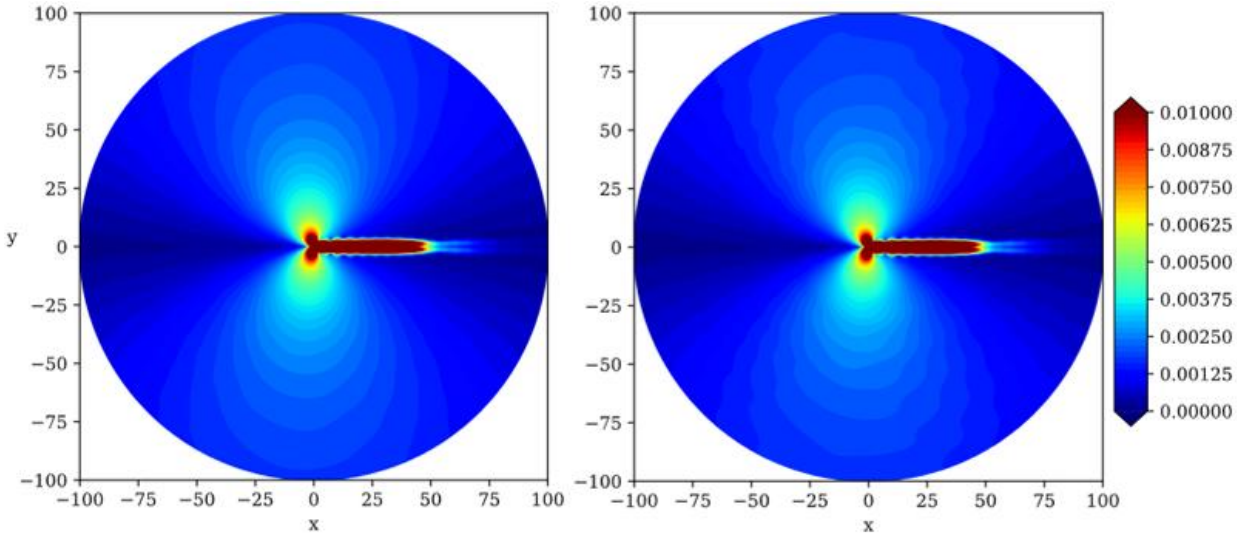


Fig. 5. Line graphs illustrating pressure fluctuations at the 100th, 110th and 120th snapshot.

3.3. Root mean square pressure

This section presents a quantitative and qualitative analysis of p' using the Root Mean Square (RMS) value, visualized through contour plots and directivity patterns.

Figure 6 presents a comparison of RMS values of p' between the actual and the predicted values. The contour plots show a close agreement in the overall distribution of p'_{rms}, indicating that the model effectively captures the primary characteristics of p' field. Nonetheless, slight differences in the contour boundaries between the actual and predicted plots are observed. These discrepancies likely stem from the variations in the p'values discussed in Figure 5, where subtle differences at certain time steps and angles are identified. While the general pattern of p' is accurately reproduced, the differences in contour boundaries suggest that the model's accuracy in capturing finer details of the pressure field could be further improved.



(a) Contour plot of actual p'_{rms}. (b) $p'_{predicted}$ plot.

Fig. 6. Contour plot of root mean square pressure (p'_{rms}).

The directivity patterns derived from the RMS values of both the actual and predicted p'provide a visual comparison that allows a detailed assessment of the prediction model's accuracy across various angular positions. The alignment between the actual and predicted patterns highlights the model's capability of capturing the directional behavior of p', as shown in Figure 7.

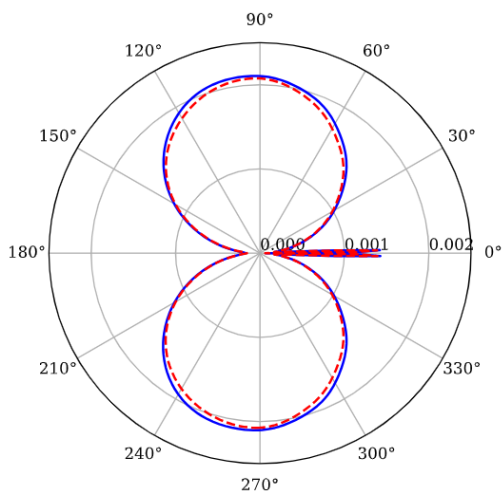


Fig. 7. Directivity pattern from RMS value of $p'(x, y, t)$; the blue line represents the actual p'_{rms} and the red dashed line represents the predicted p'_{rms}.

4. Conclusion

The linear regression ML techniques, with a learning rate of 1×10^{-2}, is employed to model the acoustic field of flow past a cylinder at a Re_∞ of 150 and a Ma_∞ of 0.2. The model effectively captures p' data, providing accurate predictions. Nonetheless, there are noticeable discrepancies in the contour plots between p'_{actual} and $p'_{predicted}$, where the predicted values show a wavy pattern. This suggests the presence of acoustic noise or other disturbances

affecting the pressure field, indicating the need for further refinement of the model to address these issues and enhance its overall accuracy.

To address these limitations, future work should prioritize enhancing the model's performance, potentially by incorporating advanced ML techniques. If the signal remains periodic with a fixed frequency, increasing the number of input features may help the model learn better representations, ultimately improving its prediction accuracy. This could be a promising direction for future research, with the potential to further refine the model's generalization capabilities and applicability to more complex flow conditions.

Nomenclature

D	diameter, m
Δr	radial grid spacing, m
E	energy, J
$\boldsymbol{f}$	external body forces, m/s^2
h_θ	hypothesis function
J	cost function
k	thermal conductivity of fluid, W/m-K
Ma	Mach number
n	number of grid points
p	Pressure, Pa
p'	instantaneous fluctuating pressure, Pa
$\bar{p}$	mean pressure, Pa
Re	Reynolds number
ρ	fluid density, kg/m^3
τ	viscous stress tensor
T	temperature, K
t	time, s
$\boldsymbol{u}$	velocity vector, m/s
X_i	i-th input feature
Y_i	i-th output value
$\emptyset$	azimuthal direction
Subscripts	
rms	root mean square
r	radial direction
θ	vector of parameters

Acknowledgement

This research was conducted with the support and resources provided by the Computational Turbulence and Aerodynamics Research Laboratory (CTAR Lab) and Chiang Mai University.

References

[1] Zhao Q, Chen S, Zhang H. Study on flow-induced noise propagation mechanism of cylinder–airfoil interference model by using large eddy simulation. AIP Advances. 2022;12(6):065310.
[2] Vinuesa R, Brunton SL. Enhancing Computational Fluid Dynamics with Machine Learning. Journal of Computational Physics. 2020;408:109248.
[3] Lighthill M. Aeroacoustic prediction based on large-eddy simulation and the Ffowcs Williams–Hawkings equation. Journal of Fluid Mechanics. 1998;45(3):245-280.
[4] Shiri FM, Perumal T, Mustapha N, Mohamed R. A comprehensive overview and comparative analysis on deep learning models. Journal of Artificial Intelligence. 2024;6(1):301-360.
[5] Yousif MZ, Yu L, Lim HC. Physics-guided deep learning for generating turbulent inflow conditions. Cambridge: Cambridge University Press; 2022.

[6] Mosavi A, Wang J. Efficiency of ML models in flow-induced noise prediction. Computational Mechanics Journal. 2021;16(4):309-322.
[7] Wang Y, Zhang H. Hybrid machine learning with physical constraints for noise prediction. Journal of Applied Fluid Mechanics. 2022;17(6):783-798.
[8] Zhou B, Karniadakis G. Towards efficient CFD through machine learning: A review. Journal of Computational Physics. 2022;454:110901.
[9] Inoue O, Hatakeyama N. Sound generation by a two-dimensional circular cylinder in a uniform flow. Journal of Fluid Mechanics. 2002;471:285–314.
[10] Nagata T, Nonomura T, Takahashi S, Fukuda K. Direct numerical simulation of subsonic, transonic and supersonic flow over an isolated sphere up to a Reynolds number of 1000. Journal of Fluid Mechanics. 2020;900:1-23.
[11] Chung TJ. Applications to Acoustics. In: Computational Fluid Dynamics. Cambridge: Cambridge University Press; 2012.
[12] Anderson JD. Modern Compressible Flow with Historical Perspective. 3rd ed. New York: McGraw-Hill; 2003.
[13] Kim J, Sandberg RD. Efficient parallel CFD methods for turbulent flow simulations on multi-core CPU/GPU clusters. Journal of Computational Physics. 2012;231(12):4355-4374.
[14] Kennedy CA, Carpenter MH, Lewis RM. Low-storage, explicit Runge-Kutta schemes for the compressible Navier-Stokes equations. Applied Numerical Mathematics. 2000;35(3):177-219.
[15] Kim J. High-order compact filters for large eddy simulation. Journal of Computational Physics. 2010;229(20):7751-7762.
[16] Ferziger JH, Perić M. Computational Methods for Fluid Dynamics. 3rd ed. Berlin: Springer; 2002.
[17] James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. 1st ed. New York: Springer; 2013.
[18] Goodfellow I, Bengio Y, Courville A. Deep Learning. 1st ed. Cambridge: MIT Press; 2016.
[19] Murphy KP. Machine Learning: A Probabilistic Perspective. 1st ed. Cambridge: MIT Press; 2012.