Estimating pollutant dispersion in a two-dimensional street canyon using physics-informed neural networks

Yichen Wang [a], Bingchao Zhang [b], Hideki Kikumoto [c]

[a]*The University of Tokyo, Meguro-Ku, Tokyo, Japan, wyc23@iis.u-tokyo.ac.jp*
[b]*The Hong Kong University of Science and Technology, Kowloon, Hong Kong, China, zhangbc@ust.hk*
[c]*The University of Tokyo, Meguro-Ku, Tokyo, Japan, kkmt@iis.u-tokyo.ac.jp*

SUMMARY

Urban flow and pollution experiments are often expensive and time-consuming, leading to sparse data collection. Therefore, efficient reconstruction approaches are needed to reduce measurement demands. This study first reconstructed the mean flow field in a two-dimensional street canyon from sparse velocity data using a physics-informed neural network (PINN). Building on this prediction, a two-step PINN framework was applied to further estimate pollutant dispersion from a line source at the bottom of the canyon. The velocity and eddy-viscosity data predicted in the first step were enforced as hard constraints in the second, improving the training efficiency and accuracy in solving the transport equation. This method successfully enabled high-resolution concentration field prediction even without any measurement data. Moreover, when a small number of sparse concentration measurements were made available (39 points within the domain), the supplementary data improved the predictive accuracy by 23% and accelerated convergence by 95%.

Keywords: *urban street canyon, sparse measurements, physics-informed neural networks, mean flow field, pollutant dispersion*

1. INTRODUCTION

Compact urban structures and high population densities have made modern city environments increasingly complex, and human safety has become a critical concern. Strong winds and air pollution significantly affect pedestrian comfort and safety, necessitating effective mitigation (Yang et al., 2020). Field measurements and wind tunnel experiments have been widely conducted to better understand these mechanisms and rapid responses. However, physical limitations and high costs often result in sparse data and certain locations remain unmeasurable. To address this challenge, super-resolution approaches are increasingly gaining attention in incomplete data reconstruction. One notable method is the physics-informed neural networks (PINNs) introduced by Raissi et al. (2019), which integrates physical laws with artificial neural networks. PINNs have demonstrated a strong potential for reconstructing high-resolution flow fields from limited data. However, owing to the computational demands and optimizer limitations for turbulent problems or cases with insufficient data, accuracy tends to decline.

In a previous study, we proposed a two-step PINN framework to reconstruct and estimate both mean flow and turbulent statistical quantities (Wang et al., 2026). The results demonstrated that separate the outputs by two models can better handle the predictions of multiple unbalanced-scale output variables. Building on this foundation, this study focuses on a more complex urban environmental problem involving both air velocity and pollution. The pollutant dispersion was estimated based on the reconstructed mean flow field without measurement data, and the acceleration and improvement in accuracy were confirmed with supplementary data.

2. METHODOLOGY

2.1. Domain design and sparse data configurations

Figure 1 presents the schematic of the two-dimensional street canyon. The dimensions of the buildings and canyons were $H \times H$, where $H = 0.12$ m. The wind flow was along the x-direction and was assumed to be cyclic. The target domain, measuring $2H \times 2H$, included one canyon and two building cubes. A line source was located at the center of the canyon bottom, releasing tracer gas at an emission rate of 100 ppm·m³/s. Black dots within the domain indicate the locations of available measurement data. The mean velocity and concentration from the large eddy-simulation (LES) dataset were used as pseudo-measurement data. This dataset serves as a proxy for a real-world wind tunnel experiment, providing noise-free data and assessing the PINN's accuracy across the domain. A total of 94 velocity data and 39 concentration data were spaced at approximately a resolution of $H/6$. More data were concentrated in regions with large velocity gradients in the shear layer and at the building corners, ensuring accurate predictions for flow field reconstruction.

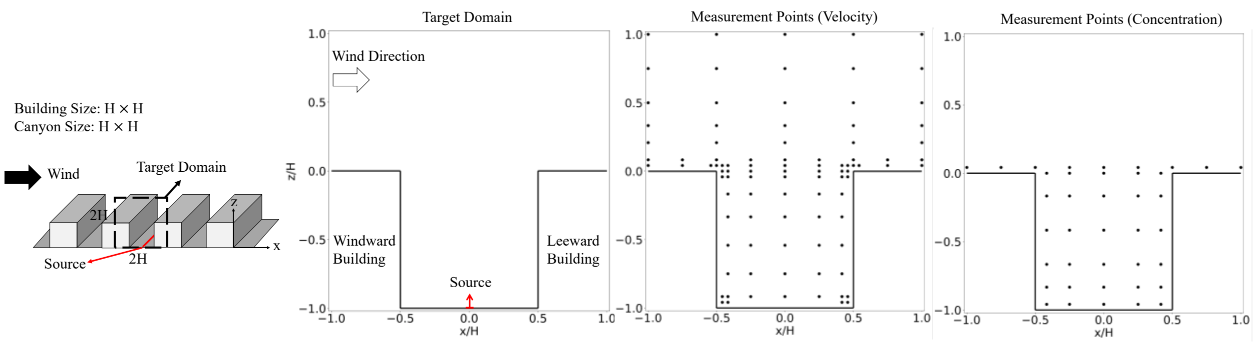

Figure 1: Street canyon model; target domain; sparse data points of velocity and concentration.

2.2. Governing equations and two-step PINN setup

The continuity (Eq. (1)) and Reynolds-averaged Navier−Stokes (RANS) equations (Eq. (2)) were used to ensure mass and momentum conservation of the turbulent flow. The pollutant dispersion is described by the transport equation (Eq. (3)). The source term was simplified and set as the concentration boundary at the line source location. Figure 2 illustrates the framework of the proposed two-step PINN model. The first step reconstructs the mean flow field. Sparse velocity data, the continuity and RANS equations, and boundary conditions were incorporated as a total loss term to train the model. The spatial coordinates vector $\boldsymbol{x}$ serve as inputs, while the outputs include the mean velocity vector $\overline{\boldsymbol{u}}$, eddy viscosity ν_t, and mean pressure $\bar{p}$. The second step focused on the concentration field. To simplify the training and improve prediction accuracy, only the concentration term $\bar{c}$ was set as the output, and only the transport equation was applied as the governing equation. The velocity and eddy viscosity obtained from step 1 were directly reused in the PDEs as constant values, thereby imposing as a hard constraint and avoiding additional training difficulties. The Reynolds number Re was calculated as 7680 using the reference velocity U_{ref} at $2H$. The Schmidt number Sc and turbulent Schmidt number Sc_t were set to 1 and 0.5, respectively.

The open-source library DeepXDE (Lu et al., 2021) was used to train the models. The first step used 30,000 steps of Adam and 20,000 steps of L-BFGS optimizer to minimize the total loss. In the second PINN, only L-BFGS optimizer is used.

$$\nabla \cdot \overline{\boldsymbol{u}} = 0 \tag{1}$$

$$\overline{\boldsymbol{u}} \cdot \nabla \overline{\boldsymbol{u}} = -\nabla \bar{p} + \nabla \cdot [(1/Re + \nu_t)\nabla \overline{\boldsymbol{u}}] \tag{2}$$

$$\overline{\boldsymbol{u}} \cdot \nabla \bar{c} = \nabla \cdot [(1/ReSc + \nu_t/Sc_t)\nabla \bar{c}] \tag{3}$$

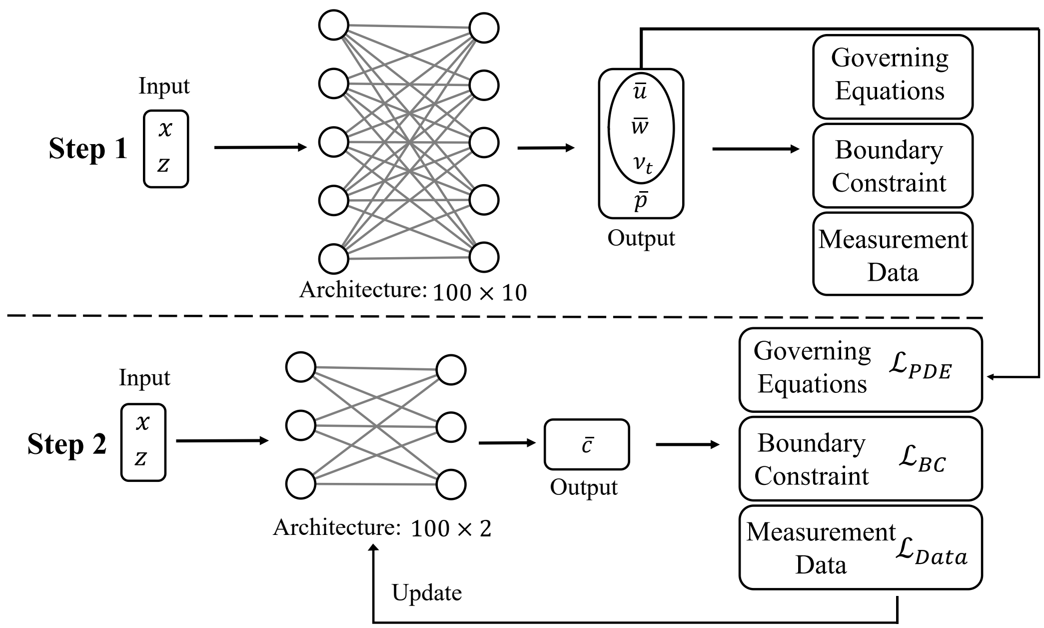

Figure 2: Framework of the two-step PINN model.

3. RESULTS

Figure 3 shows the non-dimensional reconstructed mean velocity magnitude with streamlines, ground truth, error, and eddy viscosity distributions from step 1. With the sparse velocity data, PINN successfully reconstructed the mean flow field with the key circulation vortex inside the canyon. The extra sensors at the corner and shear layer made the shear flow and counter vortices predicted. The main errors were observed near building walls, where steep velocity gradient occurred. The eddy-viscosity distribution was estimated by solving the RANS equations, for which no measurement data were provided. Negative values could be observed near the top surface of the building and in the middle of the canyon. This demonstrated that in certain areas of canyon flow, apparent counter gradient diffusion occurs owing to local turbulent eddies and complex flow.

Figure 4 illustrates the non-dimensional estimated concentration fields with and without measurement data. The results were trained in 20,000 steps for the case with no measurement data and 1,000 steps for the case with measurement data. The PINN could accurately estimate the concentration fields without requiring measurement data. The root mean square error was 0.0024 C_{ref} ($Q/U_{ref}H$, Q: emission rate). The main error was concentrated near the windward wall of the building. The large concentration gradient and the eddy viscosity predicted by the RANS equations could have led to overestimation. When supplementary data were included, the error decreased to 0.0018 C_{ref}. Additional information provides corrective guidance during training, reducing the reliance of the model on PDE constraints. Although errors remained between the data locations, the improvement was significant, reducing the number of training steps by 95 %.

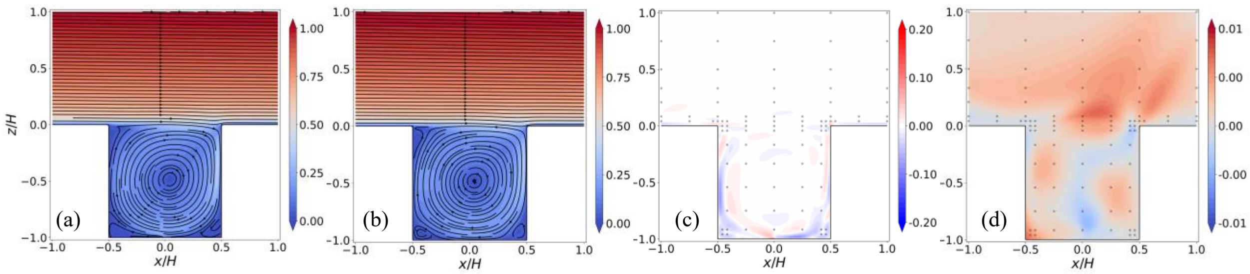

Figure 3: (a) Reconstructed mean flow field with streamlines. (b) Ground truth, (c) Prediction error (PINN - truth), and (d) Estimated eddy viscosity. The gray dots indicate the velocity data locations.

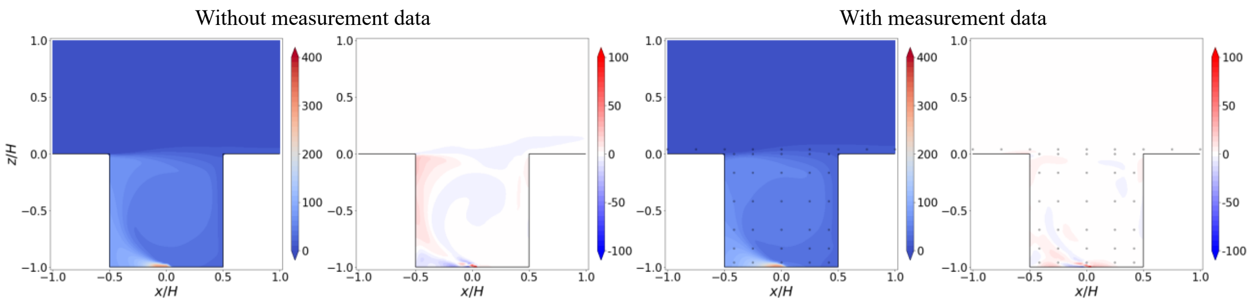

Figure 4: Estimated concentration field without (left) and with (right) concentration data with prediction error (PINN - truth). The gray dots indicate the velocity data locations.

4. CONCLUSION

In this study, we applied a two-step PINN model to reconstruct the mean flow field and estimate the pollutant dispersion in a two-dimensional street canyon. First, the proposed approach successfully reconstructed the mean flow field with high accuracy. Based on the velocity and eddy-viscosity results, the second step PINN predicted the high-resolution concentration field without any measurement data (RMSE: 0.0024 C_{ref}). Furthermore, it was also demonstrated that a small quantity of additional concentration data could enhance the prediction accuracy (RMSE: 0.0018 C_{ref}) and reduce the training step by 95%.

ACKNOWLEDGEMENTS

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 23K26257 and 24K2164.

REFERENCES

Yang, J., Shi, B., Shi, Y., Marvin, S., Zheng, Y., Xia, G., 2020. Air pollution dispersal in high density urban areas: Research on the triadic relation of wind, air pollution, and urban form. Sustainable Cities and Society 54, 101941. https://doi.org/10.1016/j.scs.2019.101941

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Wang, Y., Zhang, B., Kikumoto, H., 2026. Two-step high-resolution reconstruction of mean flow field and Reynolds stress distributions using physics-informed neural networks in a two-dimensional street canyon. Building and Environment 288, 113958. https://doi.org/10.1016/j.buildenv.2025.113958

Lu, L., Meng, X., Mao, Z., Karniadakis, G.E., 2021. DeepXDE: A Deep Learning Library for Solving Differential Equations. SIAM Rev. 63, 208–228. https://doi.org/10.1137/19M1274067