

Three-dimensional Primitive prediction with building embedded Swin Transformer for urban turbulence

Jun Park(박준),¹ Haechan Kim(김해찬),¹ Panagiotis Tsiotras,² and Sang Lee(이상봉)¹

¹*Aerospace Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea*

²*School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*

(*Corresponding author: slee1@kaist.ac.kr)

(Dated: 9 March 2026)

Large-eddy simulation (LES) is the benchmark for resolving canonical urban canopy physics critical for pedestrian comfort, pollutant dispersion modeling and Urban Air Mobility (UAM) operations. However, generating long-horizon LES data requires prohibitively long simulation times, making it impractical for real-time or large-scale applications. To overcome these limitations, we develop 3DSwinUrbanNet, a 3D Swin-Transformer-based model that directly predicts the full primitive variables (u, v, w, p) using time steps orders of magnitude larger than those in the Computational Fluid Dynamics (CFD) solver ($\Delta t = 1$ min versus LES $\Delta t = 0.05$ s). Three-dimensional (3D) building geometry embeddings are integrated into the encoder, backbone, and decoder via average-pooling over patches and linear projection into the latent space. Model accuracy is evaluated using time-averaged fields, wall-normal profiles, turbulence statistics and scale-resolved diagnostics (PDFs and energy spectra). This approach yields a 1,984 $\times$ speedup (2.58 s versus LES 5,120 s for 20 min forecasting) without precursor overhead, enabling real-time predictions on single Ray Tracing Textel eXtreme (RTX) 4090 Graphic Processing Unit (GPU).

I. INTRODUCTION

Rapid urbanization has intensified the need for accurate and fast predictions of wind flow within complex urban canopies, where building-induced turbulence strongly influences pedestrian comfort, pollutant dispersion, and emerging Urban Air Mobility (UAM) operations¹. In dense city environments, strongly sheared inflows, sharp geometric discontinuities, and multi-scale turbulent structures interact to produce highly three-dimensional (3D) and unsteady flow fields that are challenging to model and simulate reliably over practical time horizons.

The problem of wind prediction is critical for aerial vehicle navigation because environmental disturbances such as locally varying winds significantly affect the vehicle kinematics and, hence, has a huge impact on the effectiveness of the applied guidance strategies. Predicting and accounting for the prevailing wind field is particularly imperative and challenging for small-size Unmanned Aerial Vehicles (UAVs) and UAM-type aerial platforms since, owing to their small size, the wind has a large influence on their ability to maintain prescribed paths or intercept targets^{2–4}. Control strategies that anticipate the wind variations—such as fluctuations in direction or intensity—can provide more robust and predictable trajectories, leading to higher precision in reaching target sets^{4–6} or faster area coverage⁷.

In the past, level set methods have been used to solve the Hamilton-Jacobi partial differential equations in order to compute the navigation policies in time-varying flow fields^{5,6}. However, these approaches are not sufficient to solve more complex flows such as 3D turbulent structures and unsteady flow fields encountered in urban landscapes. Large-eddy simulation (LES) serves as the high fidelity benchmark for resolving urban canopy flows and associated turbulent transport,

providing detailed reference data for fundamental studies and engineering applications⁸. However, LES remains computationally prohibitive for routine design or real time forecasting, particularly when precursor domains, long physical durations, and multiple realizations over diverse building morphologies are required. These constraints become even more severe when coupling flow fields with UAM trajectory planning, which demands rapid evaluations of time evolving turbulent environments around vehicles⁹. In such settings, fast surrogate models are needed not only to accelerate individual simulations but also to enable real-time ensemble statistics, rapid scenario evaluation, and uncertainty-aware decision-making.

To address these computational challenges, data-driven approaches have leveraged recurrent neural networks, particularly Long Short-Term Memory (LSTM) networks^{10–13}, for temporal urban flow prediction. While LSTMs enable temporal predictions beyond instantaneous mapping, they remain limited by sequential processing and vanishing gradients for extended horizons. Transformer architectures offer superior temporal modeling across diverse domains^{14–17}, suggesting potential advantages for extended predictions.

Similarly, reduced-order modeling (ROM) techniques, including proper orthogonal decomposition (POD), decompose flow into modal coefficients^{10–12,18} or predict individual velocity components rather than the full primitive field (u, v, w, p). With advancing GPU capabilities, direct full-field prediction becomes feasible, as demonstrated in global weather forecasting^{19,20}. This end-to-end approach preserves coupled pressure-velocity dynamics absent in modal representations²¹.

Various architectures including convolutional neural networks (CNN)²², generative adversarial networks (GANs)²³, graph neural networks (GNNs)²⁴, and diffusion models²⁵ have been explored for urban flow prediction, typically at short

lead times with CFD-level Δt . Recent comparative studies^{26,27} demonstrate transformers’ superior turbulence prediction accuracy over CNNs and GNNs, motivating our application of the Swin Transformer^{28,29}. While GANs and diffusion models excel in super-resolution, GANs suffer training instability (mode collapse) and diffusion requires multi-step iterative inference, both precluding reliable real-time forecasting—our primary objective. This architecture excels in capturing multi-scale spatial hierarchies through shifted-window attention and hierarchical feature fusion, enabling robust long-horizon 3D urban flow predictions without modal decomposition.

However, urban environments exhibit substantial variability in building layout, height distribution, and orientation, significantly affecting flow separation, recirculation, and vertical momentum transport³⁰. While city-scale GNNs demonstrate generalization across randomly generated 2D urban layouts²⁴, 3D deep learning models for detailed LES-equivalent predictions under randomized building heights and extreme rotations remain underexplored. Such geometric robustness proves essential for resolving canonical urban canopy physics critical for pedestrian comfort and pollutant dispersion.

To overcome these limitations, we develop 3DSwinUrbanNet, a 3D Swin-Transformer-based model that directly predicts the full primitive variables (u, v, w, p) using time steps orders of magnitude larger than those in the CFD solver ($\Delta t = 1$ min vs. LES $\Delta t = 0.05$ s). Trained on LES data from urban canopy flows with randomly varied building heights and both aligned and strongly rotated configurations, the model promotes geometric robustness. Long-horizon rollout predictions up to 20 minutes assess mean-flow fidelity and turbulence statistics under coarse temporal resolution using the available test dataset.

The objectives of this work are threefold. First, 3DSwinUrbanNet’s capability to reproduce canonical urban canopy features such as leading-edge updrafts, downwash transitions, and inter-building recirculation over extended horizons is quantified through time-averaged fields, wall-normal profiles, turbulence statistics and scale-resolved diagnostics^{10,31,32}. Second, the effects of randomized building-height distributions and a 45° rotation of a fixed 4×5 array on prediction accuracy are examined to assess geometric robustness within a canonical urban canopy. Third, computational efficiency relative to LES highlights the framework’s potential as a fast surrogate for real-time urban flow forecasting, enabling practical capabilities such as rapid ensemble generation, scenario-based analyses for UAM operations, and uncertainty-aware assessment of near-canopy wind conditions.

II. METHODS

A. Overview

Fig. 1 provides an overview of the proposed framework. 3DSwinUrbanNet takes two consecutive 1-minute snapshots of the 3D velocity and pressure fields (u, v, w, p) , together with the 3D building geometry, as input and predicts the flow field at the next time step. The model was trained on LES data from

two building array configurations: Case A (buildings aligned perpendicular to streamwise direction) and Case B (45° rotation). Predictions were recursively rolled-out up to 20 min ahead. Model performance is evaluated through instantaneous fields, time-averaged flows, and turbulence statistics.

B. 3DSwinUrbanNet

3DSwinUrbanNet is constructed by modifying the global machine learning weather prediction model Aurora¹⁴. The architecture consists of an encoder-decoder structure and a 3D Swin Transformer U-Net backbone that maps the encoded representation to the flow field at the next time step²⁹. Unlike standard Vision Transformers³³, the Swin Transformer employs shifted-window self-attention²⁸ that preserves cross-window connectivity critical for capturing coherent inter-building turbulence structures across morphological discontinuities.

Both the encoder and decoder are implemented using a 3D Perceiver^{34,35}. The 3D Perceiver is one of the methods that, similar to fourier neural operators, can be trained independently of the resolution of the input grid³⁶. A randomly initialized latent array is first defined as the query, while the input data are treated as the keys and values. By computing the cross-attention between the query and the key-value pairs, the model learns a latent representation that is consistent with the input data.

The 3D Perceiver encoder in 3DSwinUrbanNet learns two latent arrays. One latent array encodes the information of the 2D building-height mesh, and the other encodes the 3D building geometry together with the flow variables u, v, w, p . Positional, scale, height, and lead-time information are embedded using a Fourier feature expansion, while the 3D building geometry is average-pooled over patches and linearly projected into the latent space.

In Aurora, the lead time is used as a conditioning variable in the backbone¹⁴. In contrast, in 3DSwinUrbanNet the 2D building height mesh is used as the conditioning field in the 3D Swin Transformer U-Net backbone. The 3D Perceiver decoder takes the embedded 3D building geometry and transforms the hidden vectors back into the physical variables u, v, w, p .

The primary difference between 3DSwinUrbanNet and Aurora lies in the utilization of building information. During the design of 3DSwinUrbanNet, injecting the full 3D building geometry alongside the flow variables u, v, w, p , and additionally embedding the building information into the encoder, backbone, and decoder, yielded the highest prediction accuracy. This configuration is therefore adopted as the definition of 3DSwinUrbanNet.

3DSwinUrbanNet is designed as a rollout prediction model, which is a commonly used strategy for sequential prediction^{14,37}. The model prediction can be written as

$$\mathcal{F}(X(t-h), X(t), C) = \tilde{X}(t+h) \quad (1)$$

where $X(t-h)$ and $X(t)$ denote the 3D fields at times $t-h$ and t , respectively, and C is a set of time-invariant context

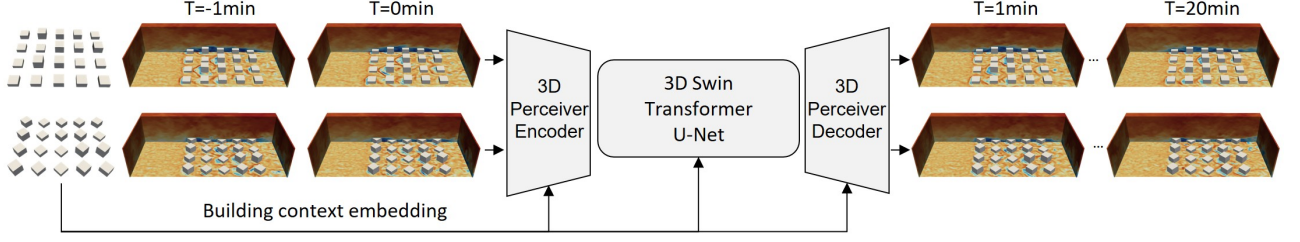


FIG. 1. Overview of 3DSwinUrbanNet

variables including the building information, position, scale, height, and lead time. The $\tilde{X}$ denotes the predicted field and $h = \Delta t$ is the temporal resolution of the data.

The rollout prediction is then performed recursively as

$$\begin{aligned} \mathcal{F}(X(t-h), X(t), C) &= \tilde{X}(t+h), \\ \mathcal{F}(X(t), \tilde{X}(t+h), C) &= \tilde{X}(t+2h), \\ &\vdots \\ \mathcal{F}(\tilde{X}(t+(i-2)h), \tilde{X}(t+(i-1)h), C) &= \tilde{X}(t+ih), \\ &\vdots \\ \mathcal{F}(\tilde{X}(t+(n-2)h), \tilde{X}(t+(n-1)h), C) &= \tilde{X}(t+nh). \end{aligned} \quad (2)$$

A well-known drawback of rollout prediction is that errors accumulate over time^{37,38}. When applied to real observational datasets that contain measurement noise and representativeness errors, this accumulation can lead to unacceptable forecast degradation, making pure rollout approaches impractical in those settings^{26,39}. However, in LES or data-assimilation datasets, where observation errors are absent or significantly reduced, rollout prediction remains feasible, and this strategy is adopted in the present study.

Additionally, a residual add method, which has shown good performance in previous machine-learning-based flow field prediction studies^{37,39–41}, was also tested. In this configuration, we added the input field $X(t)$ to the 3DSwinUrbanNet output, effectively training the model to predict the difference. This approach reduced the root mean square error (RMSE; defined in Section IV A) by approximately 50% compared to the non residual add methods. However, despite the lower RMSE, rollout predictions up to 20 min exhibited noticeable oscillations and instability in both instantaneous fields and mean profiles. Consequently, the residual add method was not adopted in the final model configuration.

C. Buildings generation

Urban flow studies have employed varied building geometries to investigate specific physical phenomena. Realistic city center morphologies have been modeled^{18,24}, alongside idealized arrays of simple building blocks^{10,22,42–44}.

Fig. 2 shows the building layouts used in this study, arranged in a 4×5 array. Height variations between adjacent

buildings induce diverse flow phenomena, including cavity-like flows³⁰. To assess 3DSwinUrbanNet's ability to predict such flows, individual building heights were randomly sampled between $H/3$ and H following Xie et al.⁴⁴, where $H = 160$ m is the reference height.

Given 3DSwinUrbanNet's windowed processing of spatial domains, buildings aligned perpendicular to the main flow direction (Case A) and rotated by 45° (Case B) represent extreme geometric variations^{45,46}. LES simulations of these configurations provided the training data to evaluate the model's sensitivity to building orientation.

In this study, the building layout is restricted to a fixed 4×5 array with uniform plan area density λ_p and inter-building spacing. Geometric variability is introduced solely through randomized building heights in $[H/3, H]$ and two orientations (aligned and 45° rotation), so that the global canopy densities remain approximately constant while local flow structures are significantly altered. The aligned and 45° -rotated configurations represent two extreme orientations of the same array relative to the mean flow direction, leading to strongly different wake interactions while remaining within a single canonical canopy family.

As a result, the present dataset samples a single aerodynamic roughness regime with limited variations in λ_p and streamwise frontal area index λ_f , and does not encompass the full diversity of urban morphologies (e.g., different array topologies, plan densities, or staggered layouts). The corresponding plan area densities, frontal area indices, and roughness parameters derived from the test dataset are reported in

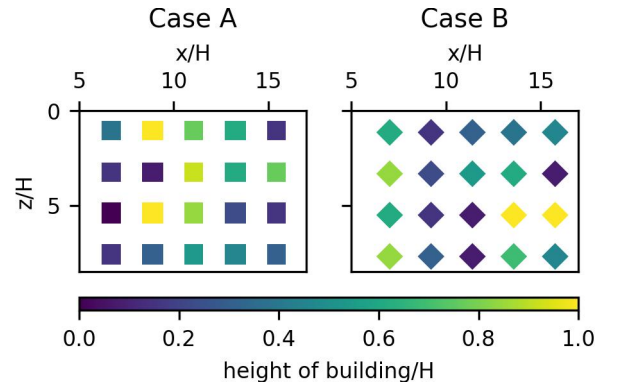


FIG. 2. Randomly generated building heights and building shapes used for learning

Section III D.

III. DATASETS AND TRAINING

A. Numerical simulation

1. Governing equations and LES methodology

The training and evaluation datasets are generated from numerical simulations of a neutral atmospheric boundary layer (ABL) using LES. In the LES framework, the large-scale turbulent motions are explicitly resolved, while the effects of unresolved subgrid-scale (SGS) motions are modeled.

The filtered incompressible Navier-Stokes equations governing the resolved flow field are written as

$$\frac{\partial \tilde{u}_i}{\partial x_i} = 0 \quad (3)$$

$$\frac{\partial \tilde{u}_i}{\partial t} + \frac{\partial}{\partial x_j} (\tilde{u}_i \tilde{u}_j) = -\frac{1}{\rho} \frac{\partial \hat{p}}{\partial x_i} - \frac{\partial \tau_{ij}}{\partial x_j} + F_i \quad (4)$$

where $\tilde{u}_i = (\tilde{u}, \tilde{v}, \tilde{w})$ denotes the spatially filtered velocity components in streamwise, wall-normal, and spanwise directions, with $x_i = (x, y, z)$. The tilde ($\tilde{\cdot}$) indicates spatial filtering at the grid scale Δ . The fluid density ρ is assumed to be constant. The SGS stress tensor is defined as $\sigma_{ij} = \widetilde{u_i u_j} - \tilde{u}_i \tilde{u}_j$, and its anisotropic component is given by $\tau_{ij} = \sigma_{ij} - \frac{1}{3} \sigma_{kk} \delta_{ij}$. The modified filtered pressure is defined as $\hat{p} = \tilde{p} + \frac{1}{3} \rho \sigma_{kk}$, where $\tilde{p}$ is the filtered pressure. The external force is decomposed as

$$F_i = F_{\text{press},i} + F_{\text{IBM},i}, \quad (5)$$

where $F_{\text{press},i}$ denotes the pressure gradient forcing and $F_{\text{IBM},i}$ represents the immersed boundary force accounting for the effects of solid building array⁴⁷.

SGS stresses are modeled using the Lagrangian Averaged Scale-Dependent (LASD) dynamic model⁴⁸ and expressed as

$$\tau_{ij}^{\text{LASD}} = -2\nu_t \tilde{S}_{ij}, \quad (6)$$

where $\tilde{S}_{ij} = \frac{1}{2}(\partial_j \tilde{u}_i + \partial_i \tilde{u}_j)$ is the filtered strain-rate tensor and ν_t is the eddy viscosity dynamically determined by the LASD formulation. Near-wall effects are treated using a logarithmic wall stress model with improved near-wall correction^{48,49}.

2. Computation domain and numerical setup

The numerical simulations are conducted using a two-domain configuration consisting of a precursor domain and a main domain, as illustrated in Fig. 3, with the instantaneous streamwise velocity magnitude at $y/H = 0.016$ shown on the bottom plane.

The computational domain is discretized using a uniform Cartesian grid with grid points $(n_{x0}, n_{x1}, n_y, n_z) =$

$(210, 576, 140, 175)$, corresponding to lengths of physical domain $(L_{x0}, L_{x1}, L_y, L_z) = (2.1, 5.76, 0.7, 1.4)$ km, where the $x0$ and $x1$ denote the streamwise directions of the precursor and main domains, respectively. This configuration results in grid spacings of $\Delta x = 10$ m, $\Delta y = 5$ m, and $\Delta z = 8$ m. The precursor domain contains approximately 5.2 million grid cells, while the main domain contains about 14.1 million cells, for a total of roughly 19.3 million cells.

Periodic boundary conditions (BCs) are applied in the lateral direction, a stress-free condition is imposed at the top boundary, and a wall-stress boundary condition is applied at the bottom surface for both the precursor and main domains. In the precursor domain, a shifted periodic boundary condition is applied in the streamwise direction. The resulting flow fields are prescribed at the inlet of the main domain. A mass-conserving outflow condition is imposed at the outlet of the main domain. The simulations are performed using a fixed time step of $\Delta t = 0.05$ s, corresponding to a maximum CFL number of 0.169.

3. In-house LES code validation

The numerical configuration largely follows that of a previous study⁵⁰, employing the same computational domain, grid configuration, and validation case. The in-house LES code is validated against the same wind-tunnel experimental data for flow around a cubical building⁵¹. Figure 4 presents comparisons of the vertical profiles of the time-averaged streamwise velocity and root-mean-square (RMS) velocity fluctuations between the LES results and experimental measurements at representative locations. The profiles are extracted at three streamwise locations relative to the cubical building, with the building leading edge defined as $x/H = 0.0$, namely at $x/H = -1.0$, $x/H = 1.5$, and $x/H = 2.5$, corresponding to the upstream, near-wake, and downstream regions, respectively. The validation focuses on the mean streamwise velocity figure 4(a) and the RMS velocity fluctuations in the wall-normal figure 4(b) and spanwise figure 4(c) directions. Overall, the LES results are in good agreement with the experimental measurements across all considered locations. Based on this validation, the present LES simulations are performed for each case to generate a physically consistent dataset used for training and evaluating the proposed machine-learning model.

B. Preprocessing

Preprocessing strategies for LES data depend on both the chosen backbone architecture and the data representation, and must therefore be tailored to each specific model configuration²³. In our case, direct training on raw LES data was constrained by GPU memory limitations. Therefore, preprocessing was performed as follows. The target domain was defined from $x/H = 0.0$ to 3.0 downstream of the last building column to focus on near-building flows, while retaining the upstream buffer region essential for sequential predictions. Within this domain, LES data were filtered using a Gaussian

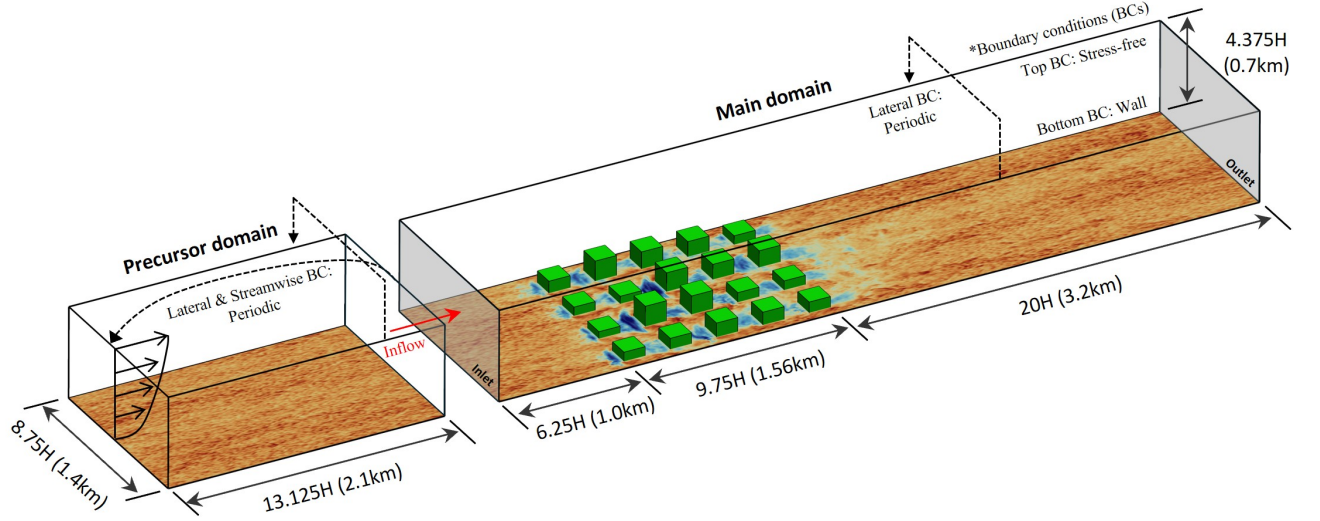


FIG. 3. Computational domains with instantaneous streamwise velocity magnitude at $y/H = 0.016$ shown on the bottom plane.

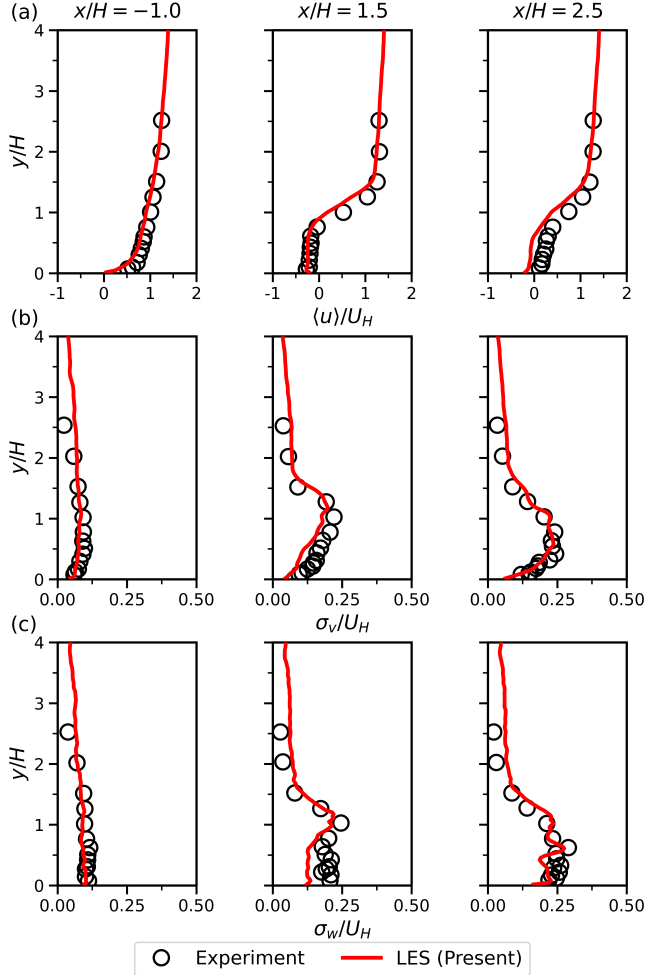


FIG. 4. Profiles of (a) the mean streamwise velocity, (b) wall-normal and (c) spanwise RMS fluctuations at $x/H = -1.0, 1.5,$ and 2.5 .

filter ($\sigma = 1.0$) and downsampled via $2 \times 2 \times 2$ block averaging.

Since 3DSwinUrbanNet employs transformer based processing with a patch size of 4, the domain was cropped such that N_x , N_y , and N_z are multiples of 4. Patch sizes of 1, 2, 4, and 8 were evaluated; sizes of 1 and 2 required excessive reductions in embedding dimensions and batch size, while size 8 yielded overly coarse resolution. Patch size 4 provided optimal predictive performance and was adopted.

C. Data division

Fig. 5 summarizes the division of the training, validation, and test sets. Data sampling from the LES simulation commenced at around 0.5 h, when the flow reached statistical convergence, with a temporal resolution of 4 s. A total of 600 samples were designated as the test set. To ensure decorrelation from the test data, 550 samples corresponding to approximately three flow-through times (FTT) were excluded⁵². The FTT is defined as L_x/U_H , where L_x is the domain length in the streamwise direction and U_H is the streamwise velocity at height H .

The remaining 7,668 samples were randomly shuffled and divided into training and validation sets, comprising 80% and 20% of the data, respectively. Additional training with 4,670 more samples was performed but showed negligible improvement; therefore, training was finalized using 7,668 samples. This procedure was applied identically for both Case A and Case B, so that the entire dataset is constructed from two LES realizations (an aligned and a 45° -rotated 4×5 canopy with randomized building heights) and thus represents a single canonical aerodynamic regime rather than a broad ensemble of distinct urban morphologies.

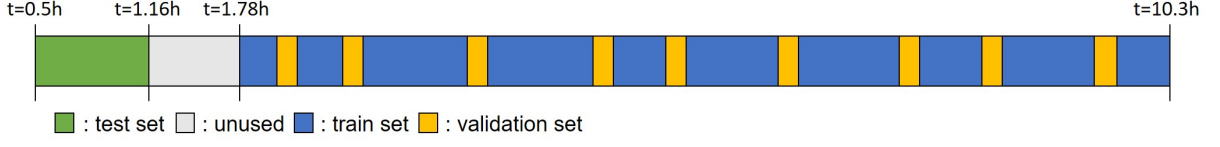


FIG. 5. Division of train, validation, and test sets. The t indicates the physical time elapsed since the simulation starts.

TABLE I. Urban canopy parameters computed from the test dataset.

Case	λ_p	λ_f	d/H	z_0/H	u_H [m s ⁻¹]	u_*/u_H
A	0.211	0.041	0.30	5.7×10^{-3}	3.91	0.09
B	0.208	0.056	0.40	4.8×10^{-3}	3.84	0.09

D. Aerodynamic canopy parameters

To characterize the aerodynamic regime of the present urban canopy, we report standard morphometric and roughness parameters computed from the test dataset. λ_p and λ_f are evaluated over a target domain that spans the 4×5 building array and is extended by half a building spacing upstream and downstream in the streamwise direction.

Within the same domain, planar- and time-averaged mean velocity profiles $U(z)$ are computed from the test dataset and used to estimate the characteristic velocity u_H at $z = H$, as well as the roughness parameters friction velocity u_* , displacement height d , and roughness length z_0 via log-law fitting,

$$U(z) = \frac{u_*}{\kappa} \ln \left(\frac{z-d}{z_0} \right), \quad (7)$$

where κ is the von Kármán constant.

The resulting parameters are summarized in Table I. For Case A, we obtain $\lambda_p = 0.211$ and $\lambda_f = 0.041$, with $u_H = 3.91$ m s⁻¹, $d/H \approx 0.30$, $z_0/H \approx 5.7 \times 10^{-3}$, and $u_*/u_H \approx 0.09$. For Case B, $\lambda_p = 0.208$ and $\lambda_f = 0.056$, with $u_H = 3.84$ m s⁻¹, $d/H \approx 0.40$, $z_0/H \approx 4.8 \times 10^{-3}$, and $u_*/u_H \approx 0.09$. All parameters are consistently derived from the test dataset to match the configuration used for model evaluation.

E. Temporal correlation function

The temporal correlation function (ρ) is an important metric when designing models and interpreting results in turbulence prediction using deep learning^{26,39,53}. It is defined as

$$\rho(\tau) = \frac{\sum (u(t) - \overline{u(t)})(u(t+\tau) - \overline{u(t+\tau)})}{\sum (u(t) - \overline{u(t)})^2}, \quad (8)$$

where the overline $\overline{(\cdot)}$ denotes the time average. ρ was computed over the test dataset, with results shown in Fig. 6. Notably, ρ decreases to approximately 0.45 when $\tau = 1$ min, justifying the 1 min temporal resolution adopted in this study.

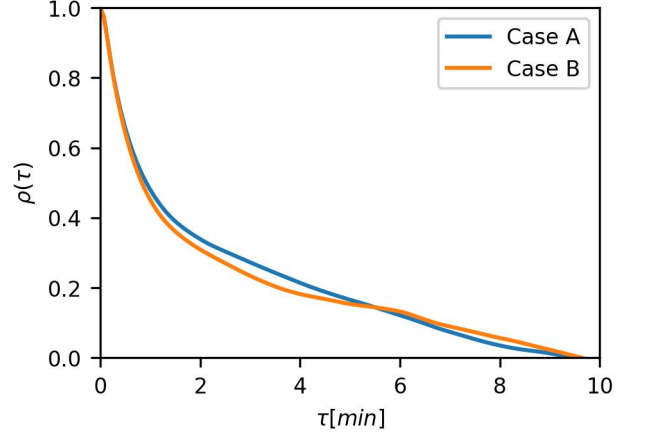


FIG. 6. Temporal correlation function ($\rho(\tau)$)

F. Model training

Model training was performed in a multi-node setting using two compute nodes, each equipped with four NVIDIA RTX 4090 GPUs. The network was trained for 100 epochs with a batch size of 2 using the Adam optimizer. The learning rate was set to 5×10^{-5} with β -parameters $(\beta_1, \beta_2) = (0.9, 0.999)$. The mean squared error (MSE; defined in Section IV A) was employed as the loss function.

Training solely on Case A data produced reasonable single-step RMSE for Case B predictions with correct building morphology. However, multi-step rollouts progressively projected Case A geometry onto Case B predictions, a phenomenon exacerbated without building embeddings. This behaviour is treated as an ablation result that highlights the necessity of explicit 3D building-geometry conditioning and joint training on both Case A and Case B to obtain stable long-horizon predictions across different orientations of the same canopy. Consequently, joint training on both Case A and Case B data was adopted to ensure robust generalization across rotated morphologies.

IV. RESULTS AND EVALUATIONS

This section presents the results and analyses of the present study. First, the evaluation metrics employed for analysis are described. Next, we evaluate temporal prediction accuracy, which is critical for long-term time-evolving predictions. Prior to turbulence characteristic analysis, instantaneous fields are visualized, followed by comparisons of time-averaged fields.

Subsequently, turbulence characteristics are analyzed, and computational times are compared with those of LES.

A. Evaluation metrics

In this study, we employ MSE and RMSE to quantify point-wise prediction errors:

$$\text{MSE}(X_j) = \frac{1}{N} \sum_{j=1}^N (X_j - \tilde{X}_j)^2 \quad (9)$$

$$\text{RMSE}(X_j) = \sqrt{\frac{1}{N} \sum_{j=1}^N (X_j - \tilde{X}_j)^2}, \quad (10)$$

where X represents filtered LES ground truth and $\tilde{X}$ denotes model predictions. RMSE values are reported in physical units for direct comparison with LES.

B. Temporal prediction

Temporal prediction in urban flow has traditionally relied on LSTM^{10–13}. Recent advances in transformers, demonstrated by their successful rollout predictions in operational global weather forecasting^{14,37}, motivate their application to noise-free filtered LES data. To determine the optimal time step for 20-min predictions, we first evaluated performance across different temporal resolutions. LES requires 24,000 iterations at $\Delta t = 0.05$ s to reach 20 min, underscoring the computational challenge. As shown in Fig. 6, ρ were 0.97, 0.85, and 0.45 for $\tau = 4$ s, 12 s, and 1 min, respectively. Higher ρ facilitates single-step prediction, but rollout performance over extended horizons requires evaluation.

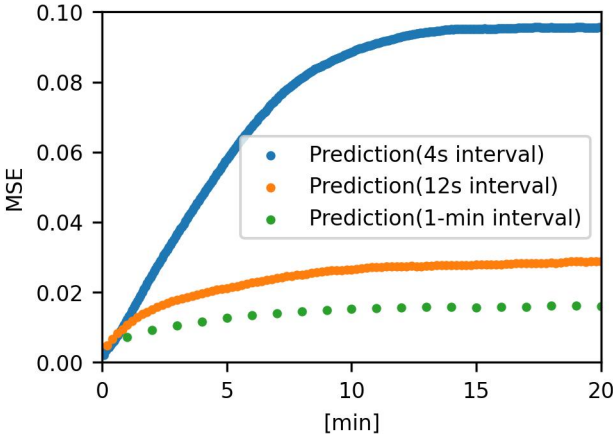


FIG. 7. Time series prediction Mean Square Error (MSE) rollout: 4-second, 12-seconds and 1-minute Intervals up to 20 Minutes

Fig. 7 shows MSE results for training with $h = 4$ s, 12 s, and 1 min per Eq. (1), rolled out to 20 min using a Swin Transformer backbone conditioned on lead time h . MSE is computed between predicted fields $\tilde{X}(t + ih)$ and test set reference $X(t + ih)$, where X comprises u, v, w, p .

For single-step predictions ($i = 1$), smaller h yields lower MSE due to higher temporal correlation ρ , demonstrating effective lead-time conditioning. As rollout steps i increase, MSE grows progressively for all h , characteristic of error accumulation. The h -dependent single-step performance motivated adoption of $h = 1$ min, after which the architecture was modified to use 3D building geometry as conditioning instead of lead time for final urban flow predictions.

Complex turbulence phenomena occur near buildings; thus, normalized RMSE was computed within the region spanning $2H$ upstream of the leading building edge, $2H$ downstream in the streamwise direction, and up to height $2H$. Temporal RMSE profiles for u, v, w , vorticity components $\omega_x, \omega_y, \omega_z$, Reynolds shear stress ($-\overline{u'v'}$), turbulent intensity (TI), and turbulent kinetic energy (TKE) are presented for Case A and Case B in Fig. 8.

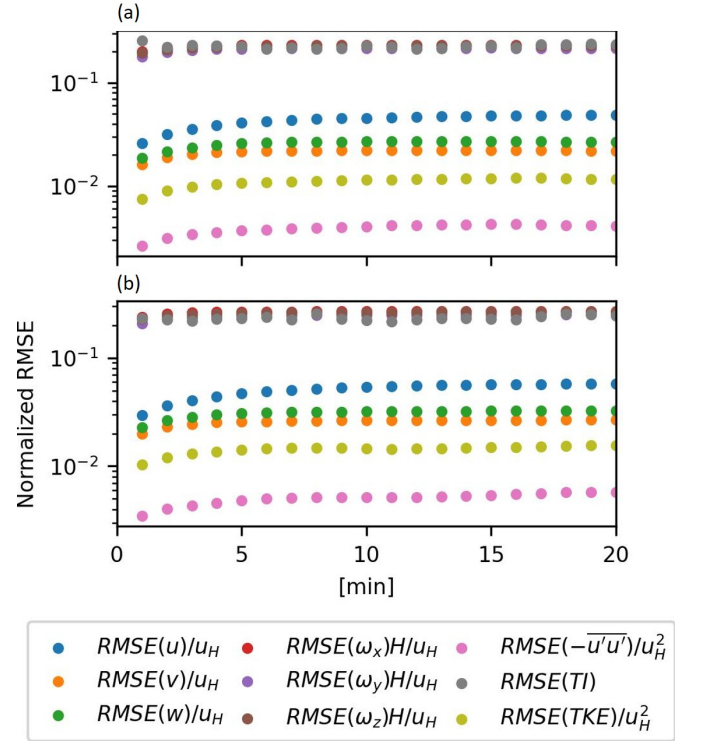


FIG. 8. Normalized RMSE results of temporal predictions for (a) Case A and (b) Case B.

To evaluate second-moment quantities, we first define the temporal averaging operator as

$$\bar{V}_i = \frac{1}{N} \sum_{t=1}^N V(t + ih), \quad i = 1, 2, \dots, 20, \quad (11)$$

where the average spans $N = 271$ test-set time instants and $V \in \{u, v, w\}$, enabling robust evaluation despite rollout error accumulation. Here, the overbar denotes a temporal mean at a given lead time index i .

Fluctuations are then defined as

$$V' = V - \bar{V} \quad (12)$$

and TI and TKE are given by

$$TI = \sqrt{\frac{u'^2 + v'^2 + w'^2}{3}} \frac{1}{U} \quad (13)$$

$$TKE = \frac{1}{2} (\overline{u'^2} + \overline{v'^2} + \overline{w'^2}) \quad (14)$$

where U denotes the local velocity magnitude computed from instantaneous u, v, w components. Normalization for RMSE was performed using the characteristic length scale H and the spatially averaged streamwise velocity u_H at height H from the LES data.

TI exhibits nearly constant RMSE over time, whereas other parameters show increasing RMSE with lead time, eventually saturating. Notably, while vorticity RMSE exhibits consistently higher absolute values than velocity components, its temporal degradation rate remains comparable to TKE and slower than velocity RMSE up to 20-min lead times. This indicates that small-scale rotational structures are preserved better than bulk velocity fields despite rollout error accumulation.

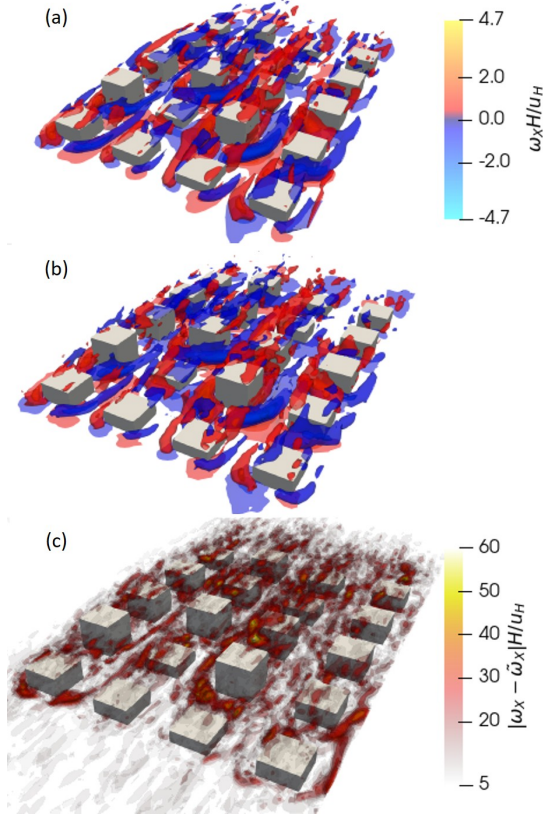


FIG. 9. Vorticity of filtered LES data and predicted data at 1-min lead time, and their difference for panels (a), (b), and (c).

Urban flow studies have established the formation of horseshoe vortices at building bases and tip vortices above buildings^{31,32}. To validate 3DSwinUrbanNet's ability to capture these coherent structures, vorticity fields at 1-min lead

time are compared in Fig. 9: (a) filtered LES data, (b) 3DSwinUrbanNet predictions, and (c) their absolute difference. The results confirm faithful reproduction of both vortex structures, consistent with the sustained vorticity RMSE performance observed in Fig. 8.

C. Instant field

Prior to turbulence statistical analysis, instantaneous field comparisons validate 3DSwinUrbanNet's capacity to capture coherent 3D turbulence structures.

Single step Q-isosurfaces in Fig. 10 reveal the model's ability to capture turbulent structures immediately downstream of building wakes, achieving qualitative agreement with filtered LES for both Case A (buildings aligned normal to streamwise direction) and Case B (45° rotation).

In both cases, fine-scale structures are under-resolved due to the model's coarse temporal resolution ($h = 1$ min), yet the model faithfully captures building geometry and height variations through learned embeddings. In particular, the progressive development of smaller-scale structures further downstream from building wakes, particularly from the second column to the fifth (rear) column in the 4×5 building array, characteristic of urban canopy turbulence evolution⁵⁴, is accurately reproduced, demonstrating geometry-aware prediction that validates building embedding efficacy and establishes fidelity for quantitative turbulence statistics.

D. Time-averaged field

Streamline patterns and vertical momentum transport vary with inter-building spacing and height variations^{10,30}.

Both filtered LES reference fields and 3DSwinUrbanNet predictions were processed *identically* using Eq. (11). While LES datasets typically require larger ensembles for higher-order statistics, the time-averaged field analysis focuses on first moment (mean velocity) fidelity of 3DSwinUrbanNet relative to filtered LES. This consistent temporal averaging across both datasets ensures fair model evaluation without data deficiency concerns for the targeted statistical comparisons.

Streamlines from side (Fig. 11) and top (Fig. 12) views demonstrate 3DSwinUrbanNet's mean flow fidelity.

In Fig. 11, comparison of filtered LES (a) and 1-min prediction (b) shows excellent agreement in: (i) upstream flow impingement ahead of the first column, (ii) cavity-like recirculation between buildings, and (iii) vertical flows induced by height differences between the fourth and fifth columns. Similarly, 20-min prediction (d) versus filtered LES (c) maintains these key features even at long lead times, confirming model robustness across vertical positions ($z/H = 1.0, 3.2, 5.3, 7.5$).

Fig. 12 further validates geometry-aware prediction at $y/H = 0.19$. At 1-min horizons, inter-building flows are faithfully reproduced for both Case A and Case B. At 20-min predictions, characteristic degradation manifests as main stream-

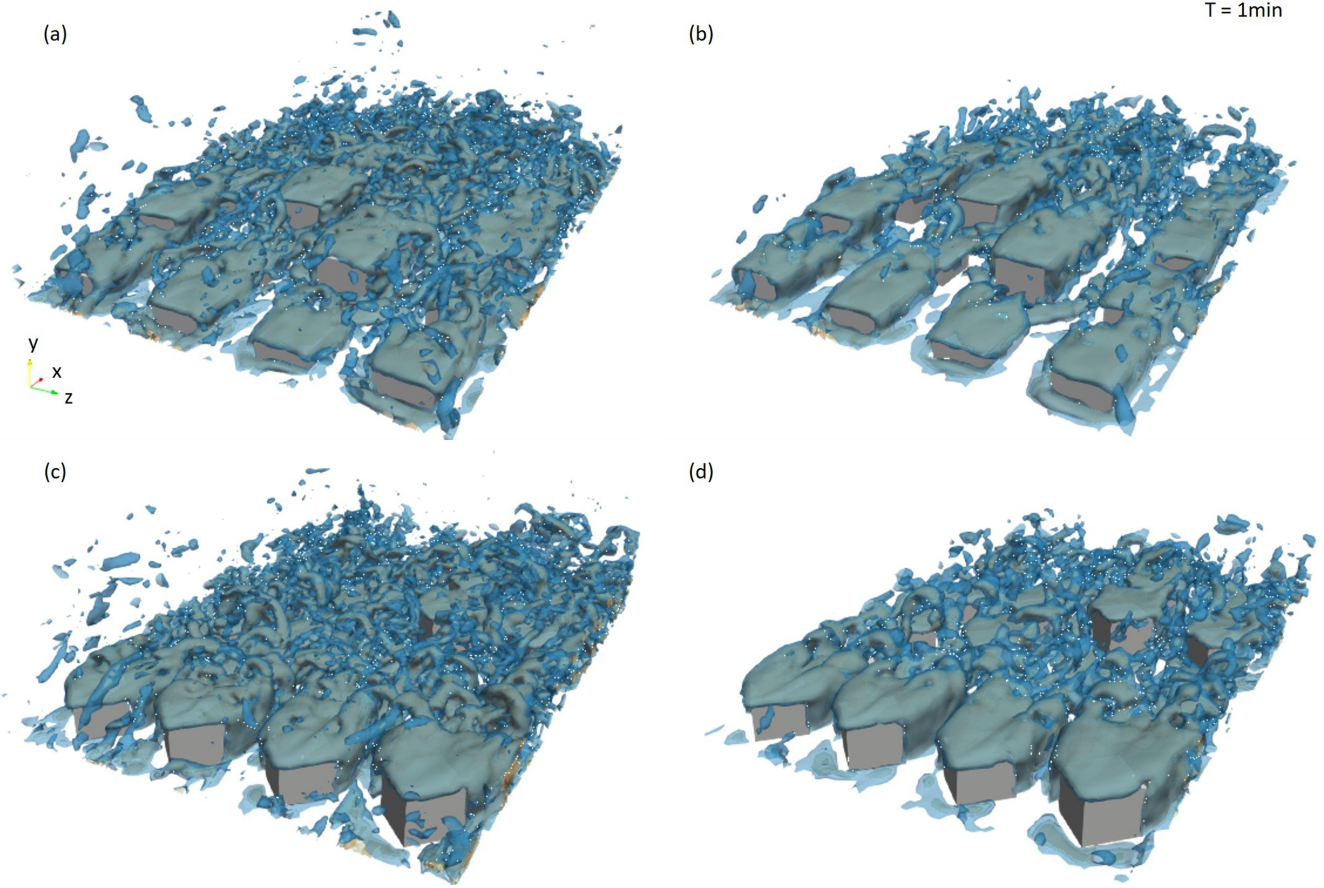


FIG. 10. Q-isosurfaces at 1-min lead time: filtered LES vs. 3DSwinUrbanNet prediction. Panels (a) and (c) show the target results, while panels (b) and (d) show the predicted results. Panels (a) and (b) correspond to Case A, and panels (c) and (d) correspond to Case B

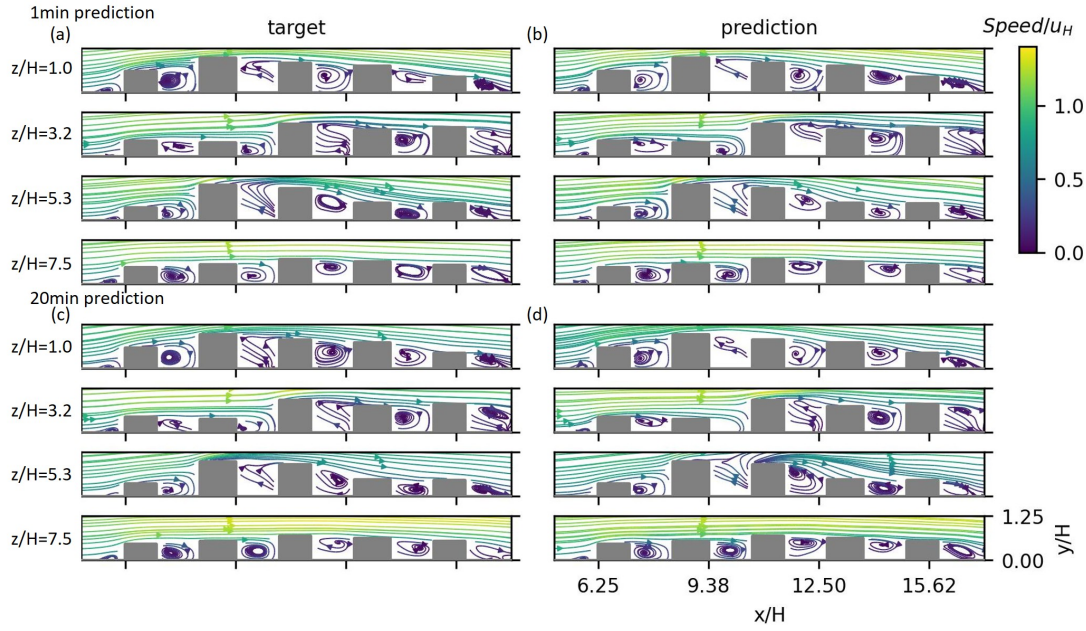


FIG. 11. Streamlines of the time-averaged target data (first column) and predicted data (second column). For a 1-minute interval, target and prediction data at four different vertical locations ($z/H = 1.0, 3.2, 5.3, 7.5$) are shown in panels (a) and (b). Target and prediction data after a 20-minute interval are shown in panels (c) and (d).

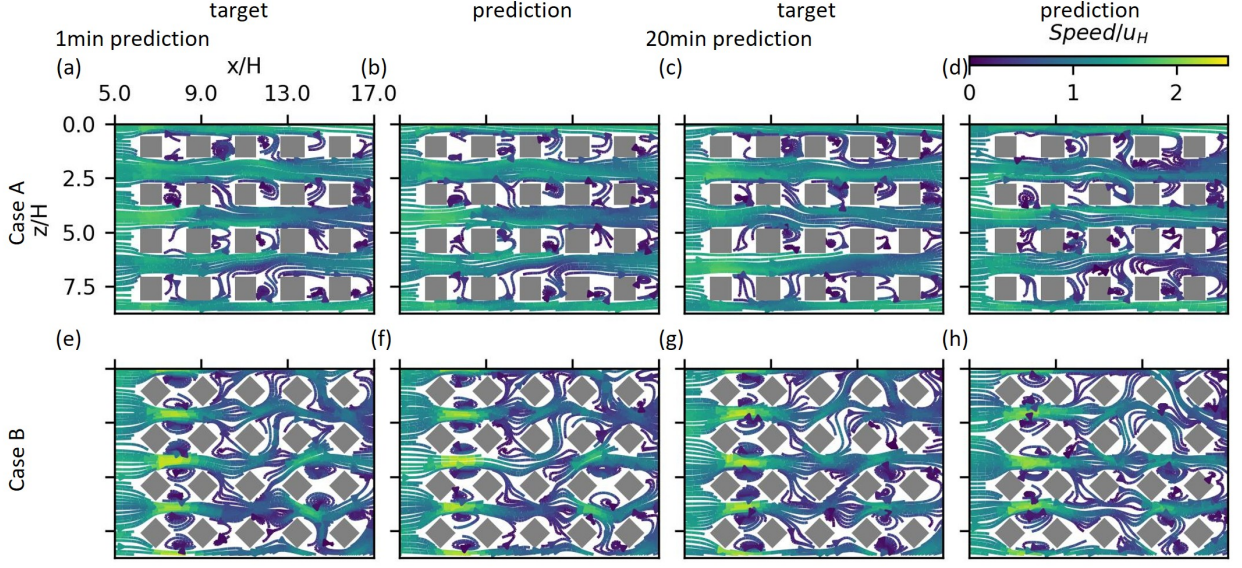


FIG. 12. Top-view streamlines of time-averaged fields at $y/H = 0.19$. First/third columns: filtered LES target; second/fourth columns: 3DSwinUrbanNet predictions. Rows (a-d): Case A; rows (e-h): Case B. Left column: 1-min lead time; right column: 20-min lead time.

wise flows being deflected by persistent inter-building secondary flows, consistent with expected error accumulation.

Wall-normal velocity profiles (Fig. 13) at $y/H = 0.1875, 0.375, 0.625$ capture canonical urban canopy flow physics: strong updrafts form along building faces at the leading edge, transitioning to downward flow for $y/H < 0.25$ ¹⁰, with substantial persistence above height H .

Inter-building vertical transport is generally well-predicted, validating 3DSwinUrbanNet’s accuracy for inlet region flows—defined as the leading-edge impingement zone where oncoming flow first encounters the building array—and essential vertical momentum transport critical for urban ventilation and pollutant dispersion modeling. However, at $y/H = 0.625$, predictions exhibit moderately higher vertical velocities compared to filtered LES for both Case A and Case B, indicative of slight overprediction in inter-building momentum exchange.

E. Turbulence statistics

Reynolds shear stress ($-\overline{u'v'}$), TI, and TKE provide stringent benchmarks for evaluating higher-order turbulence prediction fidelity. These second-moment statistics are particularly sensitive to small-scale dissipation and energy transfer processes, making them critical for validating machine learning models in complex urban flows.

Filtered LES reference statistics were computed over the full test set ($N = 600$ time instants, ~ 40 min coverage), ensuring well-converged turbulence quantities representative of statistically stationary urban canopy flows. In contrast, 3DSwinUrbanNet second-moment statistics were evaluated at discrete lead times $i = 1, 2, \dots, 20$ min using Eq. (11) with available rollout predictions. This approach enables direct quantification of rollout-induced error accumulation specifically in

higher-order turbulence moments, distinguishing model performance from first-moment mean flow fidelity.

Fig. 14 and 15 juxtapose mean velocity profiles and Reynolds shear stress at $z/H = 1.0, 3.2, 5.3, 7.5$ for Case A and Case B, with unfiltered LES as solid black reference lines. Mean profiles maintain excellent agreement across all lead times, demonstrating robust primary momentum transport prediction. Reynolds shear stress (30-fold vertical magnification) reveals subtler limitations: short term predictions align well with LES, but 20-minute rollouts exhibit pronounced degradation through turbulent stress underprediction, characteristic of second moment error accumulation.

TI profiles through building row centers (Figs. 16, 17) confirm the expected turbulence hierarchy at 1-min lead time, where peak TI regions in high shear zones between buildings match filtered LES profiles closely.

At 20-min predictions, systematic underprediction of TI magnitudes across both cases signals artificial numerical dissipation of turbulent fluctuations, characteristic of long horizon deep learning rollouts. This degradation aligns with expected energy cascade blockage in under resolved predictions, where small scale fluctuations fail to sustain against viscous like damping.

Notably, Case B (45° rotation, Figs. 17) exhibits anomalous TI amplification in weakly turbulent wake recovery regions (columns 2-4), contrasting with Case A decay. This geometry specific instability suggests enhanced vortex shedding from oblique building edges, with elevated local Reynolds shear stress beyond filtered LES reference—highlighting both model limitations and sensitivity to urban morphology variations.

TKE distributions (Figs. 18, 19) mirror this degradation pattern, with excellent short lead time accuracy giving way to progressive dissipation at 20-min horizons.

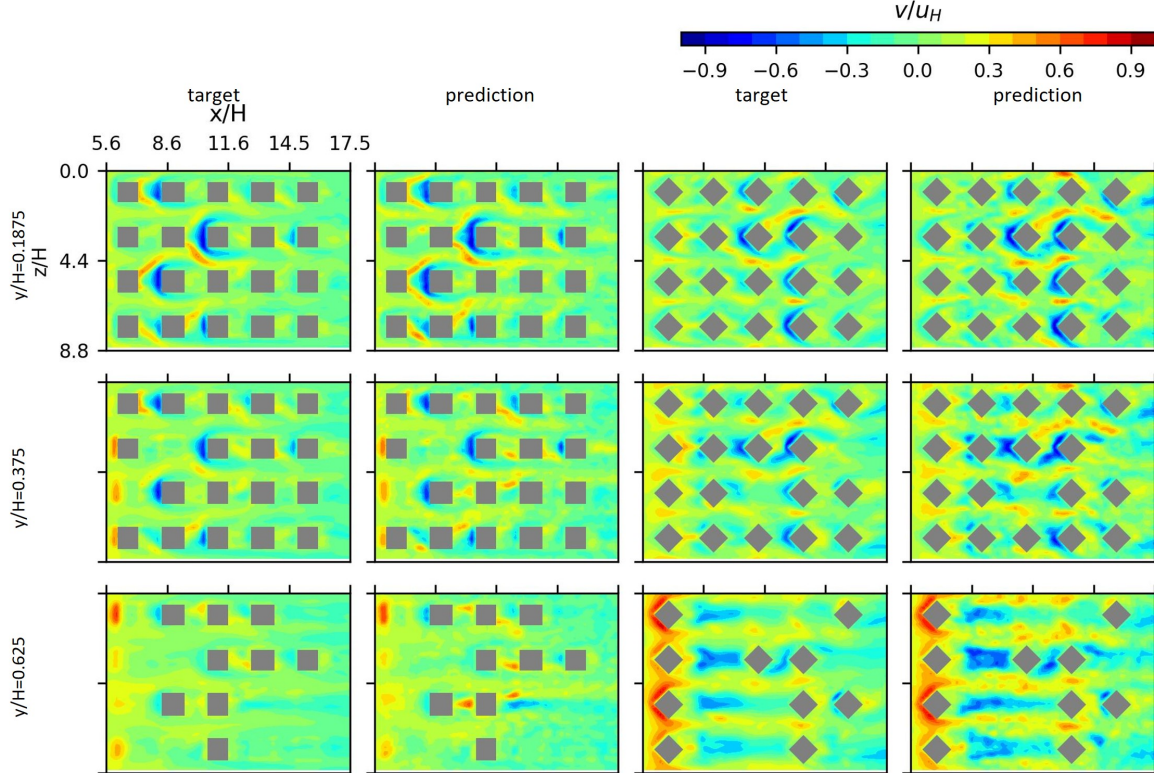


FIG. 13. Wall normal velocity profiles at various heights for the $T = 1$ min interval.

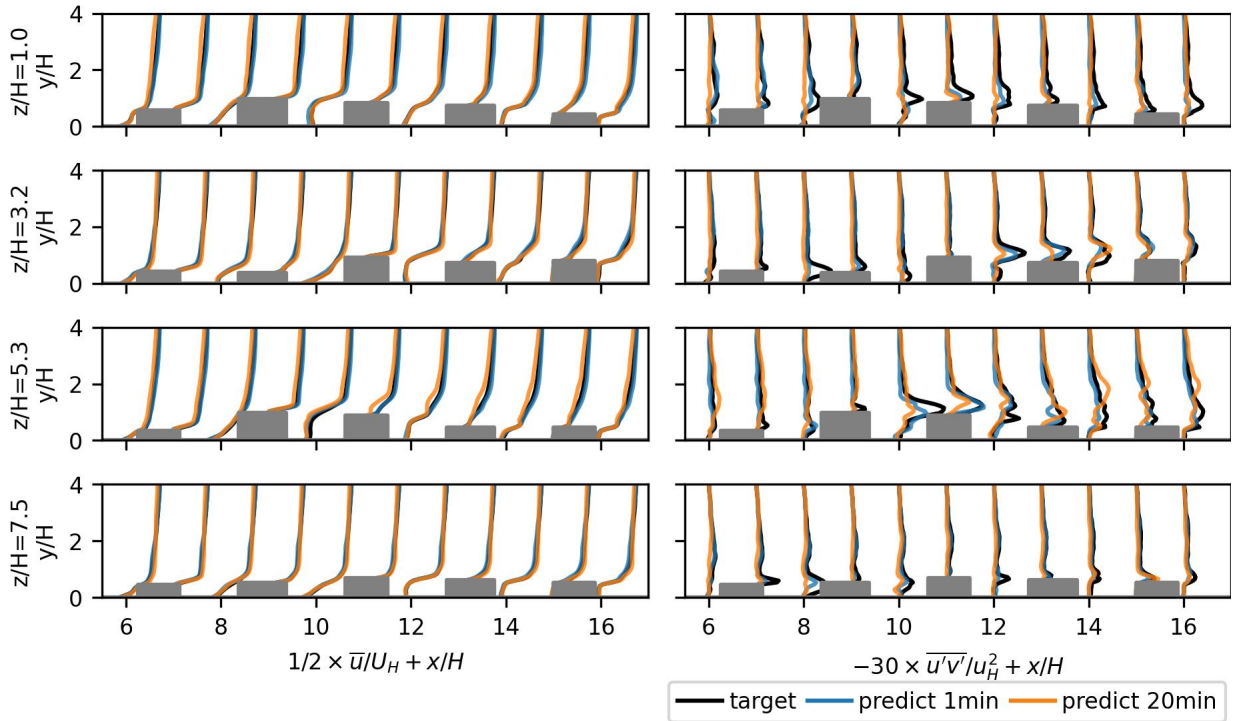


FIG. 14. Mean profile and Reynolds shear stress of Case A.

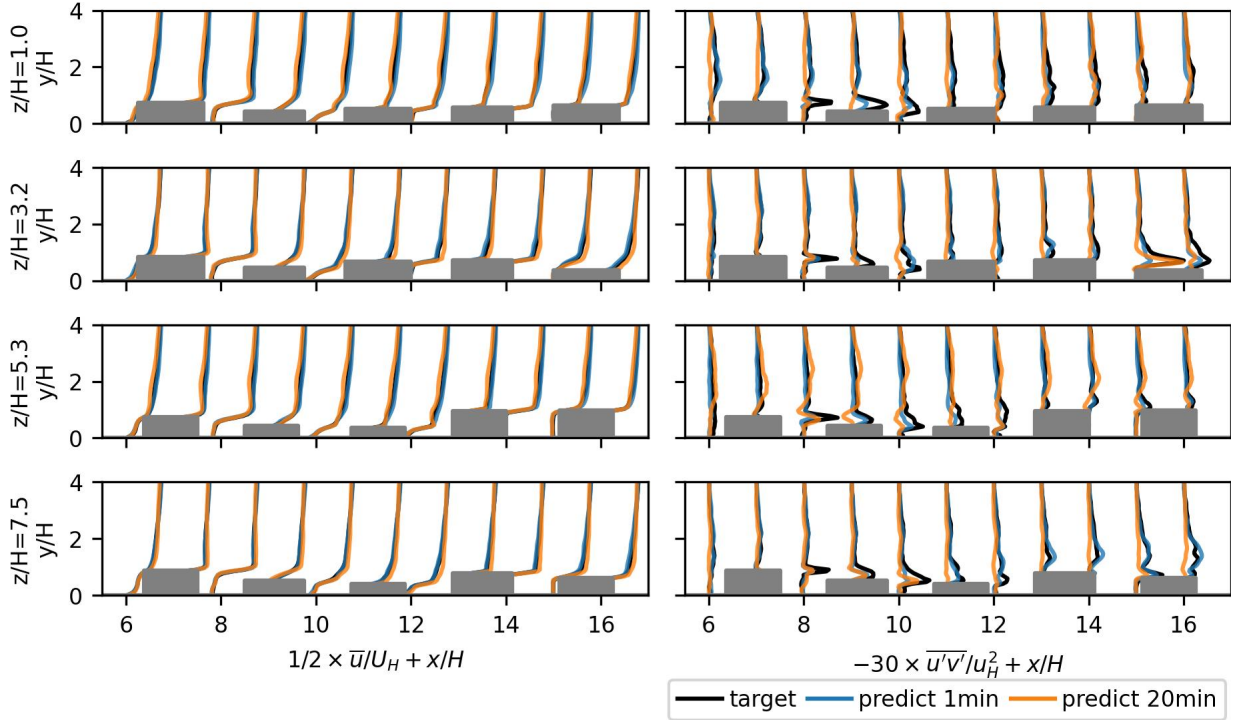


FIG. 15. Mean profile and Reynolds shear stress of Case B.

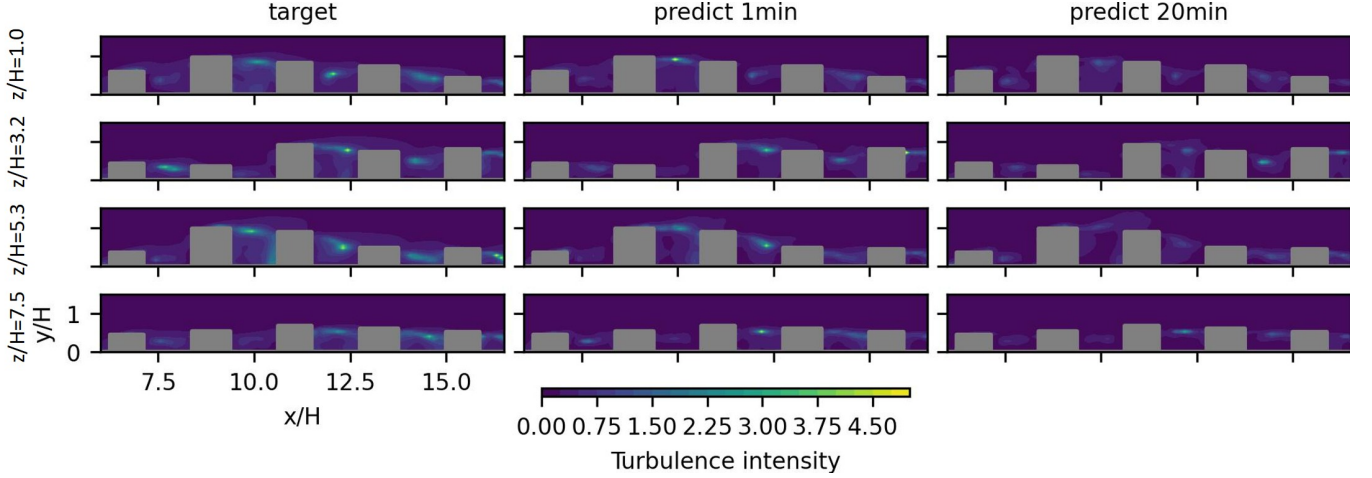


FIG. 16. Turbulent intensity (TI) of Case A. Each row corresponds to the TI at $z/H = 1.0, 3.2, 5.3$, and 7.5 . The first, second and third columns show the TI of the target data, 1-min prediction and 20-min prediction results, respectively.

In Case A (Fig. 18), TKE magnitudes consistently decay from 1-min to 20-min predictions across all building rows, reflecting systematic underprediction of turbulence production. Case B (Fig. 19) exhibits more complex behavior: while overall TKE reduction occurs, localized amplification persists between columns 4–5 at 20-min horizons, regions of enhanced inter-building shear, contrasting with the monotonic decay in Case A and highlighting geometry-specific rollout dynamics.

These results underscore 3DSwinUrbanNet’s mean flow fidelity contrasted against the inherent challenge of sustaining turbulence—particularly shear-related turbulence in oblique

configurations—during extended rollouts.

While the model is trained and evaluated on filtered LES data with randomized building-height distributions and both aligned and 45° -rotated configurations, the underlying 4×5 array maintains similar λ_p and λ_f . As quantified in Section IIID, both cases correspond to relatively open, moderately dense canopies with comparable λ_p and only moderate variations in λ_f . Consequently, the present results demonstrate geometric robustness within a single canonical canopy family, rather than generalization across a broad range of distinct urban morphologies.

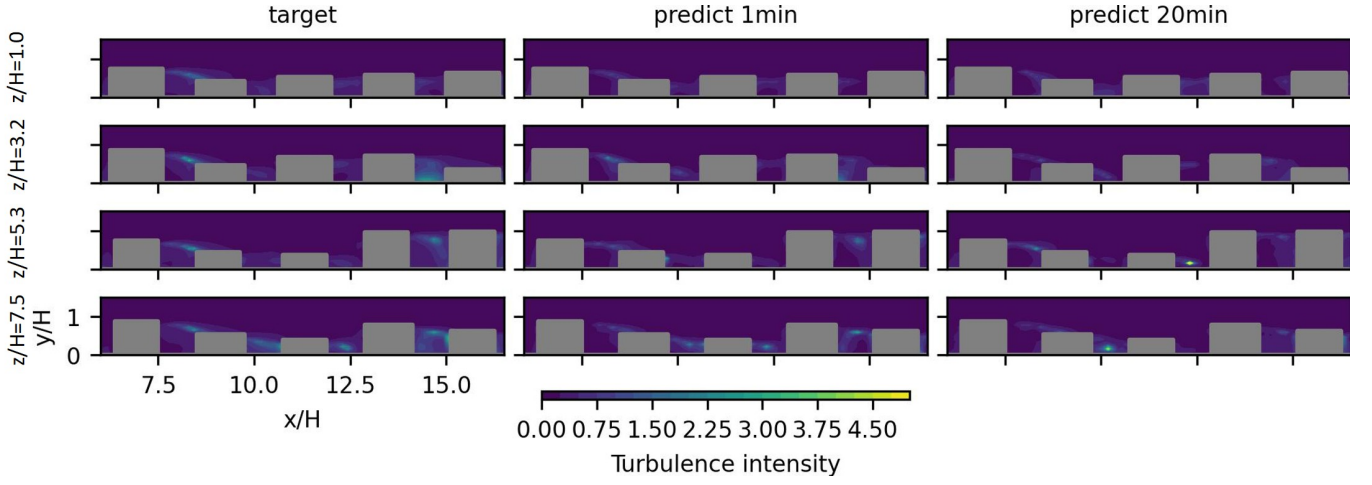


FIG. 17. TI of Case B. Each row corresponds to the TI at $z/H = 1.0, 3.2, 5.3$, and 7.5 . The first, second and third columns show the TI of the target data, 1-min prediction and 20-min prediction results, respectively.

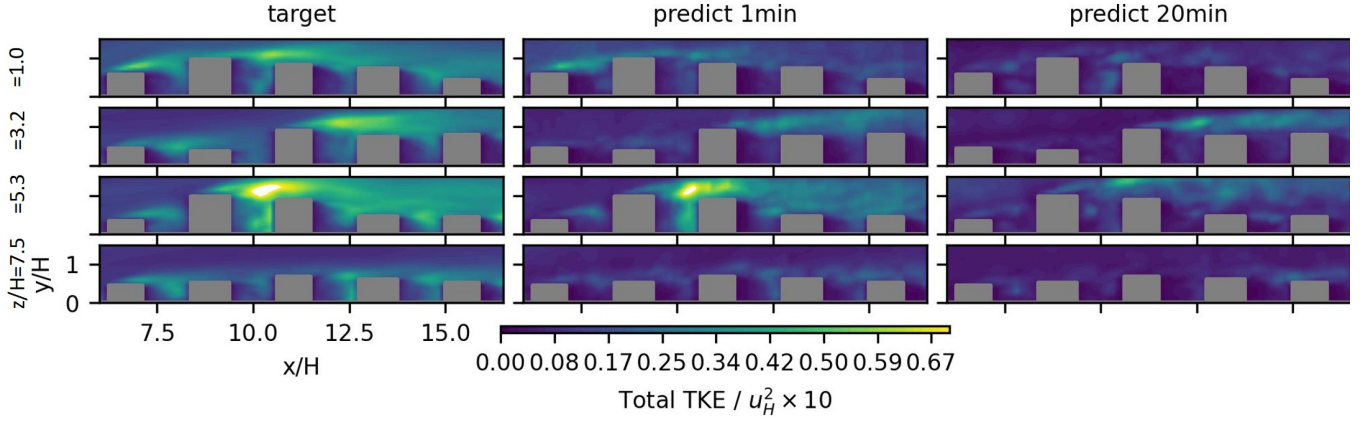


FIG. 18. Total Turbulent Kinetic Energy (TKE) of Case A. Each row corresponds to the total TKE at $z/H = 1.0, 3.2, 5.3$, and 7.5 . The first, second and third columns show the total TKE of the target data, 1-min prediction and 20-min prediction results, respectively.

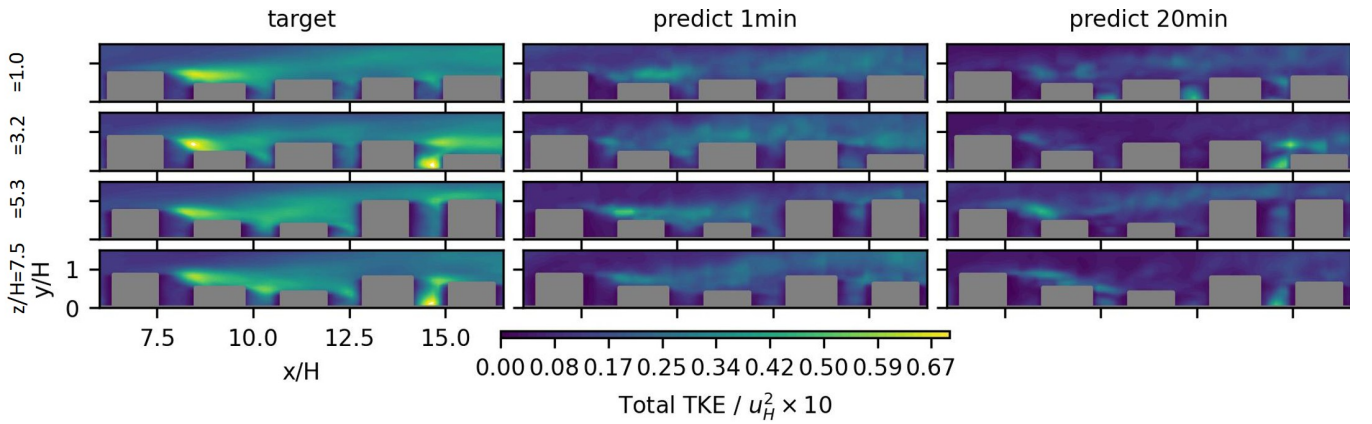


FIG. 19. TKE of Case B. Each row corresponds to the total TKE at $z/H = 1.0, 3.2, 5.3$, and 7.5 . The first, second and third columns show the total TKE of the target data, 1-min prediction and 20-min prediction results, respectively.

F. Scale-resolved diagnostics

To examine the detailed fluctuation distributions and spectral characteristics underlying the integrated turbulence quantities (TI, TKE), we analyze probability density functions (PDFs) and energy spectra.

Due to randomized building heights, PDFs and spectra are computed at $y/H = 1.0$ over the same streamwise target domain used for the canopy parameters (i.e., the region spanning the 4×5 building array, extended by half a building spacing upstream and downstream [Sec. III D]), aggregating samples over all snapshots in the test dataset.

Fig. 20 shows PDFs of the normalized streamwise velocity fluctuations u'/u_H at $y/H = 1.0$ over the canopy target domain for Case A (Fig. 20 (a)) and Case B (Fig. 20 (b)). The 1-min predictions closely match the filtered LES PDFs in both variance and shape, whereas the 20-min predictions display a broadened distribution and modified tails, indicating altered fluctuation amplitudes and intermittency at long lead times. These deviations are consistent with the trends observed in TI and TKE and further illustrate the impact of recursive rollout on higher-order statistics.

To further assess small-scale turbulence, we compute 1D energy spectra $E_{uu}(k_z)$ as spatial averages over this same canopy target domain at each snapshot, followed by temporal averaging.

Fig. 21 shows the spectra $E_{uu}(k_z)/u_H^2$ at $y = H$ in this canopy region for Case A (a) and Case B (b), plotted versus the spanwise wavenumber $k_z H$. For short lead times (1 min), the predicted spectra closely follow the filtered LES reference over low to intermediate wavenumbers up to $k_z H \approx 4$ in both cases. At 20 min lead time, the predicted spectra (orange) show slightly reduced energy levels compared to 1 min predictions across most wavenumbers, with pronounced oscillations predominantly at low-to-intermediate wavenumbers indicating increased sampling noise from rollout error accumulation. Conversely, at higher wavenumbers ($k_z H > 4$), the spectra exhibit diminished energy content with reduced oscillatory amplitude, signaling degradation in high-frequency turbulence fidelity due to artificial dissipation. These trends are consistent with the systematic underprediction of TI and TKE at extended lead times discussed in Section IV E.

G. Computational efficiency

Table II compares the computational time required for LES simulations and 3DSwinUrbanNet predictions on the same grid resolution ($N_x \times N_y \times N_z = 152 \times 68 \times 88$). LES simulations were performed on a single AMD node with 64 CPU cores. 3DSwinUrbanNet training utilized two nodes, each equipped with four NVIDIA GeForce RTX 4090 GPUs, while inference time was measured on a single RTX 4090 GPU within an Intel(R) Xeon(R) w5-3435X 16-core server.

LES requires $\Delta t = 0.05$ s with 1,200 iterations to simulate 1 minute of physical time, and 24,000 iterations for 20 minutes. The two domain method additionally requires a precursor domain with $N_x = 110$, $N_y = 68$, $N_z = 88$ grid points. In con-

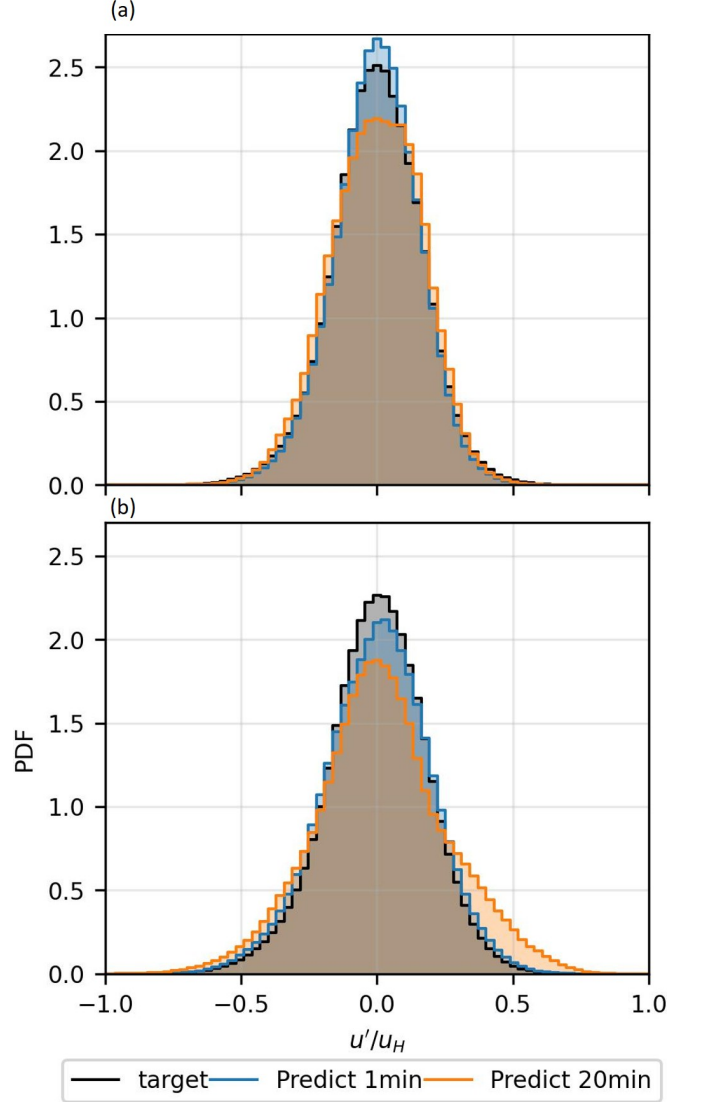


FIG. 20. Probability density functions(PDFs) of streamwise velocity fluctuation at $y/H = 1$ for Case A (a) and Case B (b).

trast, 3DSwinUrbanNet operates at $\Delta t = 1$ min, requiring only 0.9 s for a 1-minute prediction, 284 times faster than LES. 20-minute predictions take 2.58 s on a single RTX 4090 within an Intel(R) Xeon(R) w5-3435X 16-core server, achieving a 1,984-fold speedup over LES on a single AMD Ryzen threadripper 3990x 64-core node despite $4 \times$ fewer CPU cores. Even assuming maximum CPU advantage for LES, hardware contributes $< 0.2\%$ ($4/1984$) to total speedup, with algorithmic efficiency dominating.

This millisecond-level inference time on a single GPU makes it feasible to generate large ensembles of 10–20 min urban wind forecasts for different initial conditions, inflow realizations, or randomized height configurations within the present canonical canopy family at essentially negligible additional cost. In practice, this enables real-time ensemble statistics and rapid scenario-based analyses that are prohibitively expensive with LES, particularly when the flow predictions

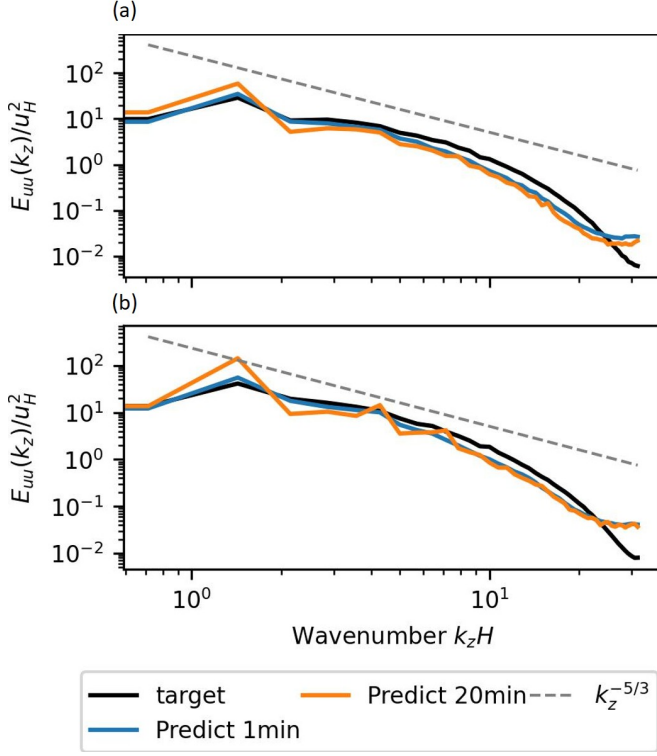


FIG. 21. One-dimensional energy spectra of the streamwise velocity fluctuations $E_{uu}(k_z)/u_H^2$ at $y = H$ in the canopy region for Case A (a) and Case B (b) as a function of spanwise wavenumber $k_z H$. Black: filtered LES; blue: 1-min prediction; orange: 20-min prediction; grey dashed: $k_z^{-5/3}$ reference slope.

TABLE II. Comparison of computational time between LES and 3DSwinUrbanNet

Method	simulation length		hardware specification
	1-min	20-min	
LES	256 s	5120 s	AMD 1node (64cores)
3DSwinUrbanNet	0.9 s	2.58 s	1 GPU (NVIDIA 4090)

are repeatedly queried as inputs to UAM trajectory planning and guidance algorithms.

V. CONCLUSIONS

3DSwinUrbanNet advances urban turbulence prediction through an end-to-end Swin Transformer architecture with geometry-aware building embeddings. Unlike standard transformers where window partitioning discards immediate neighboring information, the shifted-window mechanism²⁸ preserves cross-window connectivity essential for capturing complex inter-building dynamics across randomly varied heights and extreme 45° rotations (Case B). This enables direct primitive field (u, v, w, p) forecasting without decomposition, achieving 20-min urban canopy predictions using coarse $\Delta t = 1$ min time steps (equivalent to $1200 \times \text{LES } \Delta t = 0.05$ s)

on LES-equivalent grids ($152 \times 68 \times 88$).

The model’s exceptional mean flow fidelity—validated through time-averaged streamlines, wall-normal velocity profiles, and vertical momentum transport—reproduces canonical urban canopy physics: leading-edge updrafts, downwash transitions ($y/H < 0.25$)⁵⁴, cavity recirculations, and height-induced vertical flows persisting above H . This robustness stems from building geometry embeddings that encode morphological variations (randomized heights and a 45° rotation of a fixed 4×5 array) directly within the Swin Transformer backbone, enabling geometry-aware turbulence prediction within a single canonical canopy configuration without explicit mesh morphing or preprocessing.

Turbulence statistics reveal second-moment underprediction (TI, Reynolds shear stress, TKE) at extended rollouts, while scale-resolved diagnostics (PDFs, spectra) confirm degraded fluctuation intermittency (broadened tails) and high-frequency fidelity (reduced $k_z H > 4$ energy with low- k_z oscillations), reflecting deep learning’s challenge in sustaining small-scale production during iterative rollouts. Snapshot comparisons demonstrate clearest deficits in Reynolds shear stress (30× magnification) and geometry-specific TI/TKE dissipation in rotated configurations, while temporal RMSE confirms minimal errors at 1-min lead time, followed by monotonic increase that saturates around 5-min horizons. This hierarchy reflects deep learning’s inherent strength in first-moment pattern recognition versus the challenge of sustaining small-scale production mechanisms during iterative rollouts.

Computational efficiency provides 1,984× speedup (2.58 s vs. LES 5,120 s for 20 min) on single RTX 4090 GPUs without precursor spin-up, enabling real-time operational forecasting. Unlike traditional CFD requiring supercomputing clusters, 3DSwinUrbanNet democratizes high-fidelity urban flow prediction for real-time applications. In particular, the combination of LES-level mean-flow fidelity and millisecond-level inference makes it practical to compute real-time ensemble statistics and rapidly evaluate multiple wind scenarios for UAM operations and other urban-flow-sensitive applications.

These results are obtained for a neutral, canonical urban canopy with fixed plan area density and inter-building spacing, and randomized building heights and orientation. Generalization is therefore expected primarily for canopies that share a similar λ_p , λ_f , spacing, and array topology to the present 4×5 configuration, whereas application to markedly different morphologies should be regarded as extrapolative in the current form. Extension to more diverse urban morphologies with varying λ_p , λ_f , spacing, and array topologies is left for future work, and will likely require additional LES data and/or transfer-learning strategies across different canopy parameter regimes.

Future work will integrate physics-informed loss terms targeting second moment fidelity and multiscale turbulence closure (e.g., spectral constraints) for sustained rollout accuracy. The predicted turbulent flow fields will be utilized as relevant input environments for UAM trajectory planning and guidance algorithms, supporting more robust and predictable trajectories and improved precision under wind variations including fluctuation in direction and intensity. In particular,

rapid wind flow generation will allow the efficient on-line computation of higher-order statistics of the flow, that can be utilized in recent distributional control techniques that are able to directly handle the dispersion of the trajectories for small-size aerial vehicles^{55,56}.

ACKNOWLEDGMENTS

This study was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA), Republic of Korea grant funded by the Ministry of Land, Infrastructure and Transport [Grant Number RS-2022-00143965]

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Jun Park: Conceptualization (lead); Data curation (lead); Investigation (lead); Methodology (lead); Project administration (equal); Resources (lead); Software (lead); Supervision (lead); Validation (lead); Visualization (lead); Writing – original draft (lead). **Haechan Kim:** Data curation (supporting); Methodology (equal); Resources (equal); Writing – review & editing (supporting). **Panagiotis Tsiotras:** Project administration (supporting); Supervision (supporting); Writing – review & editing (supporting). **Sang Lee:** Conceptualization (equal); Data curation (supporting); Funding acquisition (lead); Project administration (equal); Supervision (lead); Writing – review & editing (supporting).

DATA AVAILABILITY STATEMENT

The data supporting this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹R. Mori, “Fuel-saving climb procedure by reduced thrust near top of climb,” *Journal of Aircraft* **57**, 800–806 (2020).
- ²E. Bakolas and P. Tsiotras, “Minimum-time paths for a light aircraft in the presence of regionally-varying strong winds,” in *AIAA Infotech@Aerospace* (Atlanta, GA, April 20–22, 2010) AIAA Paper 2010-3380.
- ³E. Bakolas and P. Tsiotras, “Feedback navigation in an uncertain flow-field and connections with pursuit strategies,” *AIAA Journal of Guidance, Control, and Dynamics* **35**, 1268–1279 (2012).
- ⁴R. P. Anderson, E. Bakolas, D. Milutinović, and P. Tsiotras, “Optimal feedback guidance of a small aerial vehicle in a stochastic wind,” *Journal of Guidance, Control, and Dynamics* **36**, 975–985 (2013).
- ⁵W. Sun, P. Tsiotras, and A. Yezzi, “Multiplayer pursuit evasion games in 3-dimensional flow fields,” *Dynamic Games and Applications* (2019), (to appear).
- ⁶W. Sun, P. Tsiotras, T. Lolla, D. Subramani, and P. Lermusiaux, “Multiple-pursuer-one-evader pursuit evasion game in dynamic flow fields,” *AIAA Journal of Guidance, Control, and Dynamics* **40**, 1627–1637 (2017).
- ⁷E. Bakolas and P. Tsiotras, “Optimal partitioning for spatiotemporal coverage in a drift field,” *Automatica* **49**, 2064–2073 (2013).
- ⁸N. Nazarian, E. S. Krayenhoff, and A. Martilli, “A one-dimensional model of turbulent flow through “urban” canopies (mlucm v2. 0): updates based on large-eddy simulation,” *Geoscientific Model Development* **13**, 937–953 (2020).
- ⁹Y. Dai, F. Zhu, W. Tu, H. Zhu, D. Qin, H. Wang, and Z. J. Zhai, “Urban air mobility in the built environment: A review of aerodynamic interactions, thermal effects, and simulation challenges,” *Sustainable Cities and Society* , 106853 (2025).
- ¹⁰Y. Lee and S. Lee, “A prediction of urban boundary layer using recurrent neural network and reduced order modeling,” *Building and Environment* **276**, 112804 (2025).
- ¹¹P. A. Srinivasan, L. Guastoni, H. Azizpour, P. Schlatter, and R. Vinuesa, “Predictions of turbulent shear flows using deep neural networks,” *Physical Review Fluids* **4**, 054603 (2019).
- ¹²G. Borrelli, L. Guastoni, H. Eivazi, P. Schlatter, and R. Vinuesa, “Predicting the temporal dynamics of turbulent channels through deep learning,” *International Journal of Heat and Fluid Flow* **96**, 109010 (2022).
- ¹³R. Vinuesa, “Perspectives on predicting and controlling turbulent flows through deep learning,” *Physics of Fluids* **36** (2024).
- ¹⁴C. Bodnar, W. P. Bruinsma, A. Lucic, M. Stanley, A. Allen, J. Brandstetter, P. Garvan, M. Riechert, J. A. Weyn, H. Dong, *et al.*, “A foundation model for the earth system,” *Nature* , 1–8 (2025).
- ¹⁵Y. Nie, “A time series is worth 64 words: Long-term forecasting with transformers,” *arXiv preprint arXiv:2211.14730* (2022).
- ¹⁶H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35 (2021) pp. 11106–11115.
- ¹⁷H. Wu, J. Xu, J. Wang, and M. Long, “Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting,” *Advances in neural information processing systems* **34**, 22419–22430 (2021).
- ¹⁸D. Xiao, C. Heaney, L. Mottet, F. Fang, W. Lin, I. Navon, Y. Guo, O. Matar, A. Robins, and C. Pain, “A reduced order model for turbulent flows in the urban environment using machine learning,” *Building and Environment* **148**, 323–337 (2019).
- ¹⁹J. Pathak, S. Subramanian, P. Harrington, S. Raja, A. Chattopadhyay, M. Mardani, T. Kurth, D. Hall, Z. Li, K. Azizzadenesheli, *et al.*, “Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators,” *arXiv preprint arXiv:2202.11214* (2022).
- ²⁰K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, “Accurate medium-range global weather forecasting with 3d neural networks,” *Nature* **619**, 533–538 (2023).
- ²¹A. Caiazzo, T. Ilescu, V. John, and S. Schischlowa, “A numerical investigation of velocity–pressure reduced order models for incompressible flows,” *Journal of Computational Physics* **259**, 598–616 (2014).
- ²²R. Zhao, S. Zhong, and R. You, “Application of convolutional neural network for efficient turbulence modeling in urban wind field simulation,” *Physics of Fluids* **36** (2024).
- ²³G. Calzolari and W. Liu, “Accelerating large eddy simulations of urban airflow with generative adversarial networks,” *Building and Environment* , 113622 (2025).
- ²⁴Z. Liu, S. Zhang, X. Shao, and Z. Wu, “Accurate and efficient urban wind prediction at city-scale with memory-scalable graph neural network,” *Sustainable Cities and Society* **99**, 104935 (2023).
- ²⁵P. Du, M. H. Parikh, X. Fan, X.-Y. Liu, and J.-X. Wang, “Conditional neural field latent diffusion model for generating spatiotemporal turbulence,” *Nature Communications* **15**, 10416 (2024).
- ²⁶J. Park and C. Lee, “Tptnet: A data-driven temperature prediction model based on turbulent potential temperature,” *Earth and Space Science* **11**, e2024EA003523 (2024).
- ²⁷M. Nawaz, B. Ali, and T. Karim, “Advancing data-driven turbulence modeling with transformer neural networks: A comprehensive review,” *Next*

- Research, 100618 (2025).
- ²⁸Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision* (2021) pp. 10012–10022.
 - ²⁹Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, *et al.*, "Swin transformer v2: Scaling up capacity and resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022) pp. 12009–12019.
 - ³⁰G. Zampino, M. Atzori, and R. Vinuesa, "Turbulence around two obstacles in tandem: Effects of obstacle height and separation," *Physics of Fluids* **37** (2025).
 - ³¹H. F. Wang and Y. Zhou, "The finite-length square cylinder near wake," *Journal of Fluid Mechanics* **638**, 453–490 (2009).
 - ³²G. Zampino, M. Atzori, E. Zea, E. Otero, and R. Vinuesa, "Aspect-ratio effect on the wake of a wall-mounted square cylinder immersed in a turbulent boundary layer," *International Journal of Heat and Fluid Flow* **112**, 109672 (2025).
 - ³³A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929 (2020).
 - ³⁴A. Jagle, F. Gimeno, A. Brock, O. Vinyals, A. Zisserman, and J. Carreira, "Perceiver: General perception with iterative attention," in *International conference on machine learning* (PMLR, 2021) pp. 4651–4664.
 - ³⁵J.-B. Alayrac, J. Donahue, P. Luc, A. Miech, I. Barr, Y. Hasson, K. Lenc, A. Mensch, K. Millican, M. Reynolds, *et al.*, "Flamingo: a visual language model for few-shot learning," *Advances in neural information processing systems* **35**, 23716–23736 (2022).
 - ³⁶Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," arXiv preprint arXiv:2010.08895 (2020).
 - ³⁷R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, *et al.*, "Learning skillful medium-range global weather forecasting," *Science* **382**, 1416–1421 (2023).
 - ³⁸T. Nguyen, R. Shah, H. Bansal, T. Arcomano, R. Maulik, R. Kotamarthi, I. Foster, S. Madireddy, and A. Grover, "Scaling transformer neural networks for skillful and reliable medium-range weather forecasting," *Advances in Neural Information Processing Systems* **37**, 68740–68771 (2024).
 - ³⁹J. Park and C. Lee, "Cprecnet: Enhanced nowcast of high-resolution short-term precipitation using deep learning," *Geophysical Research Letters* **52**, e2024GL113907 (2025).
 - ⁴⁰Y. Hu, L. Chen, Z. Wang, and H. Li, "Swinvmn: A data-driven ensemble forecasting model via learned distribution perturbation," *Journal of Advances in Modeling Earth Systems* **15**, e2022MS003211 (2023).
 - ⁴¹R. Keisler, "Forecasting global weather with graph neural networks," arXiv preprint arXiv:2202.07575 (2022).
 - ⁴²W. Liu, Y. Zou, and X. Li, "Study of interscale interactions for turbulence over the obstacle arrays from a machine learning perspective," *Physics of Fluids* **35** (2023).
 - ⁴³H. Yeo and S. Lee, "Impact of heterogeneous building arrangement on local turbulence escalation," *Building and Environment* **236**, 110217 (2023).
 - ⁴⁴Z.-T. Xie, O. Coceal, and I. P. Castro, "Large-eddy simulation of flows over random urban-like obstacles," *Boundary-layer meteorology* **129**, 1–23 (2008).
 - ⁴⁵V. Fuka, Z.-T. Xie, I. P. Castro, P. Hayden, M. Carpentieri, and A. G. Robins, "Scalar fluxes near a tall building in an aligned array of rectangular buildings," *Boundary-layer meteorology* **167**, 53–76 (2018).
 - ⁴⁶H. Foroutan, W. Tang, D. Heist, S. Perry, L. Brouwer, and E. Monbureau, "Numerical analysis of pollutant dispersion around elongated buildings: an embedded large eddy simulation approach," *Atmospheric environment* **187**, 117–130 (2018).
 - ⁴⁷G. Yao, *Immersed Boundary Method for CFD: Focusing on Its Implementation* (CreateSpace Independent Publishing Platform, 2016).
 - ⁴⁸E. Bou-Zeid, C. Meneveau, and M. B. Parlange, "A scale-dependent lagrangian dynamic model for large eddy simulation of complex turbulent flows," *Physics of Fluids* **17**, 025105 (2005).
 - ⁴⁹S. Feng, X. Zheng, R. Hu, and P. Wang, "Large eddy simulation of high-Reynolds-number atmospheric boundary layer flow with improved near-wall correction," *Applied Mathematics and Mechanics* **41**, 33–50 (2020).
 - ⁵⁰H. Yeo and S. Lee, "Impact of heterogeneous building arrangement on local turbulence escalation," *Building and Environment* **236**, 110217 (2023).
 - ⁵¹K. Sada and A. Sato, "Numerical calculation of flow and stack-gas concentration fluctuation around a cubical building," *Atmospheric Environment* **36**, 5527–5534 (2002).
 - ⁵²H. Kim, J. Kim, and C. Lee, "Interpretable deep learning for prediction of prandtl number effect in turbulent heat transfer," *Journal of Fluid Mechanics* **955**, A14 (2023).
 - ⁵³J. Kim, J. Kim, and C. Lee, "Prediction and control of two-dimensional decaying turbulence using generative adversarial networks," *Journal of Fluid Mechanics* **981**, A19 (2024).
 - ⁵⁴D. Hertwig, H. L. Gough, S. Grimmond, J. F. Barlow, C. W. Kent, W. E. Lin, A. G. Robins, and P. Hayden, "Wake characteristics of tall buildings in a realistic urban canopy," *Boundary-Layer Meteorology* **172**, 239–270 (2019).
 - ⁵⁵G. Rapakoulas and P. Tsiotras, "Stochastic control of uavs: An optimal tradeoff between performance, flight smoothness and control effort," arXiv preprint arXiv:2409.10369 (2025).
 - ⁵⁶F. Liu and P. Tsiotras, "Optimal covariance steering for continuous-time linear stochastic systems with additive martingale noise," *IEEE Transactions on Automatic Control* **69**, 2591–2597 (2024).