buildings

MDPI

Article

Rapid Pedestrian-Level Wind Prediction Across Urban Building Layouts Using a Geometry-Conditioned Multi-Scale Neural Operator

Yuanzheng Zhang and Yuecheng Yu *

School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang 212003, China
* Correspondence: 15152151761@163.com

Abstract

Convolutional and neural-operator surrogates enable rapid urban wind prediction, but their apparent accuracy can depend on physical-scale inconsistencies, related-layout leakage, and the mixture of idealized and realistic cases. The scientific question is whether a geometry-conditioned local–spectral model provides a reproducible improvement once these evaluation risks are controlled, and whether that improvement persists in realistic neighborhoods. We combine an audited UrbanTALES representation, grouped evaluation, and Geo-MSO to predict two-dimensional pedestrian-level velocity statistics and turbulent kinetic energy. The 538 cases comprise 224 idealized and 314 realistic configurations; inputs use fixed physical windows and targets are normalized by friction velocity. On 81 geometry-held-out cases, Geo-MSO achieved a wind-speed MAE of 0.1275 ± 0.0005 m s^{-1}, with baseline-relative mean reductions of 2.15% versus U-Net and 6.62% versus FNO. However, among the 47 realistic test cases, it outperformed FNO in only 51.1%, with a median FNO-minus-Geo-MSO difference of approximately 0.00010 m s^{-1}. Aggregate improvement therefore does not establish consistent realistic-domain superiority. All eight primary error diagnostics favored Geo-MSO over FNO across three seeds, whereas most U-Net paired intervals included zero. Unequal parameter counts and conditioning also prevent attribution to a single architectural component. Idealized pretraining accelerated target convergence without improving final realistic accuracy; narrow street-canyon jets remained underpredicted. Measured end-to-end inference took 0.114–1.747 s on an RTX 4090 D. The empirical contribution is a reproducible, morphology-qualified assessment of a combined model-and-evaluation pipeline, supporting preliminary screening rather than peak-wind certification or replacement of validated CFD.

Keywords: urban wind environment; pedestrian-level wind; urban microclimate; building morphology; neural operator; large-eddy simulation; domain generalization; surrogate modeling

Check for updates

Academic Editor: Haifeng Liao

Received: 15 September 2026
Revised: 9 October 2026
Accepted: 20 October 2026
Published: 24 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Urban building form changes pedestrian wind exposure, neighborhood ventilation, and the transport of heat and pollutants. CFD supplies spatially resolved evidence, but accuracy depends on the domain, forcing, turbulence model, grid, and numerical validation rather than on computational speed alone [1–3]. Recent reviews of deep learning in built-environment CFD and surrogate-assisted urban wind design place fast predictors within a broader workflow of simulation, design exploration, and verification [4,5]. The practical

problem is therefore to reduce the cost of preliminary layout and forcing comparisons without interpreting a low-cost surrogate as a certified engineering assessment.

A major existing approach represents building heights and flow fields as paired images. U-Net provides an encoder–decoder with skip connections for this representation [6]. FastFlow explores CNN-based urban velocity prediction [7], while Clemente et al. study pedestrian-level prediction in simplified urban environments with Pareto analysis of accuracy and computational cost [8]. Wang et al. evaluate a deep-learning surrogate in urban microclimate design [9]. These studies motivate raster-based screening and careful cost reporting; they do not imply that predictions retain narrow high-speed jets under every realistic morphology. Nor can latency numbers be compared fairly when field size, preprocessing, hardware, and timing boundaries differ.

Recent work has moved beyond simplified layouts. Lu et al. predict pedestrian wind from morphology and prevailing direction [10]. Chen et al. examine FNO generalization across wind directions and cities [11], Snaiki et al. use a hierarchical urban-wind model [12], and Huang et al. investigate inpainting U-Net for seamless pedestrian-field prediction [13]. City transfer, direction transfer, and domain reconstruction are thus already active research topics, not novel solely because they are included here. The present comparison instead emphasizes auditable case identities, physical-scale harmonization, grouping of related geometries, and the explicit separation of Idealized and Realistic results within a common frozen test.

Operator learning provides another family of representations. FNO learns Fourier-domain spatial mappings [14], the neural-operator framework formalizes mappings between function spaces [15], and DeepONet supplies a complementary operator-learning formulation [16]. Local–spectral hybridization itself is established: U-FNO combines convolutional and spectral processing for multiphase flow [17]. GINO addresses large-scale three-dimensional PDEs through geometry-aware representations [18]. Geo-MSO therefore does not claim the first convolution–Fourier hybrid or a new operator-approximation theorem. Its design is a task-specific, fixed-grid combination of coarse spectral mixing, local convolution, and continuous forcing modulation [19], assessed within the present pedestrian-level benchmark.

Other learning tasks require different evidence. The physics-informed graph-assisted autoencoder of Gao et al. reconstructs urban wind using sparse sensor information [20], unlike inference from geometry and forcing without flow measurements. Physics-informed machine learning and PINO use physical equations or residual constraints to regularize learning [21,22]. In the available pedestrian-height slices, absent vertical velocity and vertical derivatives prevent the evaluation of a complete incompressibility residual. Our gradient term is therefore a spatial regularizer, not a three-dimensional conservation constraint. Likewise, heteroscedastic output scales [23] require separate calibration and error-ranking tests; conformal calibration is a possible future route under appropriate exchangeability assumptions [24], not a capability demonstrated here.

UrbanTALES offers idealized and realistic LES configurations for testing these issues [25]. Its release semantics matter: native grid spacings differ; FLUX identifies pressure-gradient forcing, not a direction category; Val-* denotes resolution-validation pairs, not a machine-learning validation split; and some directory identifiers require audited mapping to metadata. The downloaded cases supply pedestrian-level statistics but not a uniform full three-dimensional training volume. A model can otherwise learn inconsistent physical scales or be evaluated on related layouts assigned to different partitions.

Three gaps motivate the scientific question. First, model comparisons need a common physical representation and independent-layout protocol to distinguish architectural behavior from preprocessing and leakage effects. Second, a mixed-domain average may

conceal realistic-neighborhood failures; family-specific errors and paired case differences are needed in addition to global metrics. Third, scalar accuracy, vector fidelity, local peak behavior, uncertainty, and deployment cost need not improve together. These are gaps in empirical evidence and evaluation design, not claims that existing urban-wind theory or surrogate methods are absent.

The aim is to determine whether the combined Geo-MSO and scale-consistent evaluation pipeline offers a reproducible, practically useful improvement under these constraints. The objectives are to (i) audit and harmonize all 538 cases and freeze case maps and grouped splits; (ii) compare Geo-MSO with U-Net and FNO using shared inputs, targets, optimization, reconstruction, and metric definitions; (iii) quantify family-specific performance, auxiliary city/direction behavior, component trade-offs, transfer, and uncertainty; and (iv) report reproducible full-field inference costs and failure-informed screening guidance. The principal contribution is this combination of data processing, evaluation protocol, and task-specific model; performance differences between unequal-capacity, differently conditioned networks are not treated as isolated component effects.

The intended application is preliminary pedestrian-level wind screening for building-layout exploration. The predictions are two-dimensional LES-derived mean statistics, not gust climatology, a full three-dimensional solution, or an independently validated comfort/safety classification. No CFD acceleration factor is claimed because matched CFD wall-clock measurements are unavailable. Final design and locations sensitive to localized high winds require validated CFD, wind-tunnel testing, or suitable field evidence.

2. Materials and Methods

We first audit case identities, geometry, grid spacing, forcing, and field availability; harmonize physical windows and nondimensional targets; and freeze grouped layouts, checkpoints, and reconstruction settings. We then train three complete implementations under a common optimizer and objective and evaluate complete fields in physical units. Analyzed factors are model implementation, dataset family, held-out geometry/city/direction population, component removal, and source-versus-scratch initialization. Comparison criteria are eight primary scalar/vector/local/structural errors, additional component and direction metrics, interval calibration, training cost, and measured end-to-end latency. Family tables are re-tabulations of frozen tests; ablations and transfer use validation data only.

2.1. UrbanTALES Data Audit and Scope

The local UrbanTALES copy contains 538 cases occupying approximately 6.05 GiB: 224 Idealized Building Blocks cases and 314 Realistic Urban Neighborhoods cases. The realistic set covers 73 cities in 22 countries. Each case includes a building-height raster and the associated pedestrian-level statistics, vertical profiles, diagnostic time series, and PALM configuration information. The fields used in this study are horizontal velocity components uped and vped, wind-speed magnitude Uped, and pedestrian-level turbulent kinetic energy TKEped. Tuwped was audited but not used as a prediction target. Pressure is not present in the local per-case release. The underlying simulations use the PALM model system described by Maronga et al. [25,26].

The dataset publication describes spatial averages, two-dimensional fields, and selected three-dimensional products [25]. However, the locally downloaded case archive did not contain a consistent three-dimensional volume field for all cases. Except for limited idealized repeated-unit information, additional three-dimensional fields require separate access. The learning problem is therefore explicitly defined on pedestrian-level two-dimensional statistics at approximately z = 1.75 m.

Directory and metadata identifiers were joined only through a frozen case map. Of 538 cases, 390 matched exactly, 35 matched after case folding, 110 were resolved one-to-one from physical signatures including resolution, domain extent, wind forcing, configuration, and morphology statistics, and 3 used aliases confirmed from the official portal. The 110 signature matches are audited engineering inferences rather than an official alias list. Raw directories remain the file-access keys; no runtime abbreviation rules are used.

`Val-*` is interpreted as a resolution-validation pair at 0.5 m spacing and is grouped with its corresponding base geometry. It is never treated as the validation split for model selection. Likewise, `FLUX` denotes a pressure-gradient-forced case. Its flow-to angle is derived from the negative horizontal pressure gradient, using +x as zero and counterclockwise angles as positive. `FLUX` is not encoded as a discrete class. Figure 1 summarizes the distributions of direction and forcing labels, mean building height, plan-area density, and raster dimensions in the two dataset families. Figure 2 illustrates representative Idealized and Realistic building geometries using native height rasters and three-dimensional model perspectives.

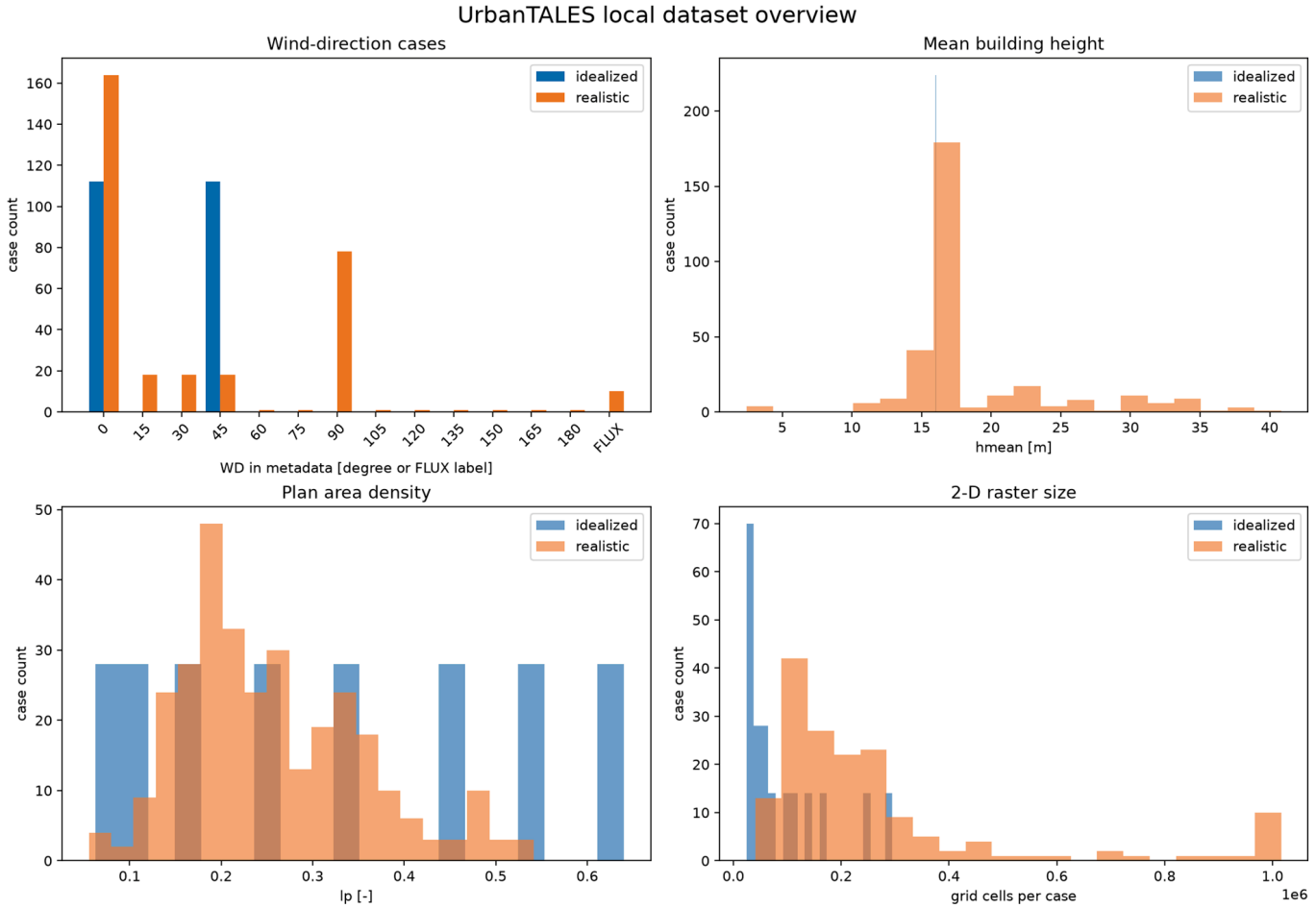


Figure 1. Distributions of characteristics in the audited UrbanTALES release: case direction/forcing labels, mean building height, plan-area density, and two-dimensional raster size, separated by dataset family. This figure summarizes dataset characteristics; it does not display representative pedestrian-level flow fields.

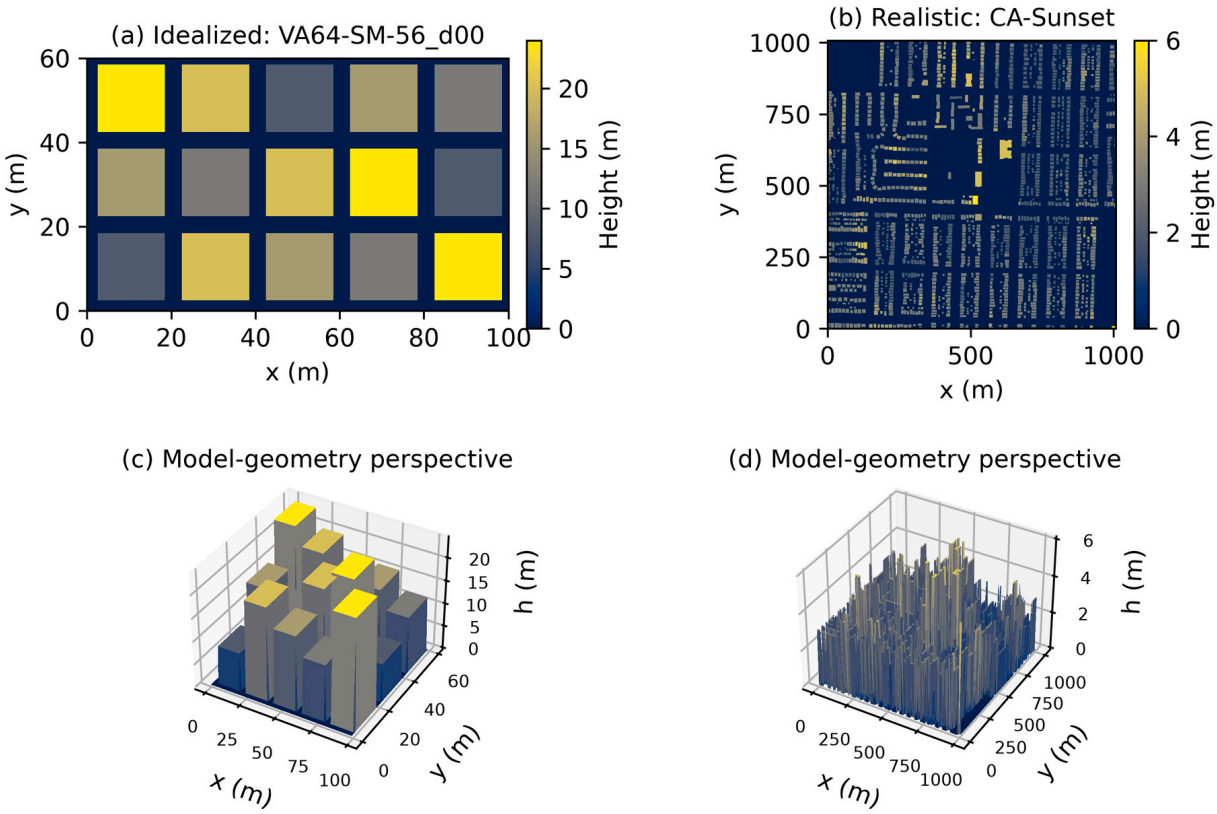


Figure 2. Modeled study geometries: (**a**,**c**) VA64-SM-56_d00 Idealized configuration and (**b**,**d**) CA-Sunset Realistic configuration. Top panels show native height rasters and bottom panels show model-geometry perspectives with coordinates/height in meters. Native spacing is read from each PALM name list (0.5 and 1.0 m, respectively). The perspectives use display-only surface subsampling; full native input extents are shown before pedestrian-field crop/resampling. They are not photographs or independent geographic surveys.

2.2. Scale-Consistent Inputs and Targets

Native cases use grid spacings of 0.5 or 1.0 m and vary substantially in spatial extent. Training samples use a fixed 256 m × 256 m physical window mapped to a 256 × 256 model grid. A 1.0 m case is sampled one-to-one, whereas each pair of native pixels in a 0.5 m case maps to one model pixel after scale-consistent resampling. Coordinates refer to cell centers in meters. Invalid or non-air pixels are excluded by target-specific masks; missing values are never replaced and scored as physical zeros.

The six input channels are normalized building height $h/50$ m, binary occupancy, signed distance to the building boundary $d/64$ m, $\cos\theta$, $\sin\theta$, and friction velocity $u\tau/0.21$ m s^{-1}. Direction and friction-velocity channels are spatially constant within a sample, but they are also provided to feature-wise linear modulation (FiLM) layers [19]. Signed distance is positive in the fluid region and retains local proximity information that occupancy alone cannot represent.

Velocity targets are nondimensionalized as $u/u\tau$, $v/u\tau$, and $U/u\tau$; turbulent kinetic energy is normalized as $k/u\tau^2$. Metrics are evaluated after inverse transformation to SI units. This avoids directly combining m s^{-1} and m^2 s^{-2} losses and makes the same network applicable across inlet forcing magnitudes. Uped is predicted as a separate LES-provided statistic rather than forced to equal $\mathrm{sqrt}(u^2 + v^2)$ because averaging operations need not commute. Table 1 summarizes the scale-consistent input and target variables, their model representations, and their physical interpretations.

Table 1. Scale-consistent model variables and transformations.

Role	Variable	Model Representation	Physical Interpretation
Input	Building height	$h/50$ m	Rasterized obstacle height
Input	Occupancy	0/1	Solid/fluid indicator
Input	Signed distance	$d/64$ m	Boundary proximity; positive in fluid
Input	Wind direction	$\cos\theta$, $\sin\theta$	Flow-to angle; +x, counterclockwise
Input	Friction velocity	$u_\tau/0.21$ m s^{-1}	Forcing magnitude
Target	uped, vped, Uped	u/u_τ, v/u_τ, U/u_τ	Pedestrian-level velocity statistics
Target	TKEped	k/u_τ^2	Pedestrian-level turbulent kinetic energy

2.3. Leakage-Controlled Partitions

Four deterministic partitions were created before formal model comparison. The primary `geometry_grouped` split groups cases by dataset family and `site_key`; the same building layout under different directions and each `Val-*`/base pair remain in one partition. `city_grouped` assigns all cases from the same country and city together. `wind_held_out` preserves layout grouping while withholding directions. The transfer partition uses grouped idealized source data and city-grouped realistic target data. An ungrouped case-level split is retained only as a diagnostic reference and is not used for the main claim. Table 2 summarizes the frozen partition sizes, grouping constraints, and evaluation purposes.

Table 2. Frozen split protocols. Counts are train/validation/test cases.

Protocol	Counts	Grouping Constraint	Purpose
iid_case_reference	377/81/80	No grouping; diagnostic only	Optimistic random-case reference
geometry_grouped	376/81/81	Family + site_key; directions and Val/base together	Primary unseen-layout test
city_grouped	220/47/47	Country + city kept together	Unseen-city test
wind_held_out	252/47/149	Grouped layouts with held-out directions	Unseen-direction test
transfer	I 176/24/24 R 220/47/47	Grouped source City-grouped target	Cross-family transfer

Model and hyperparameter choices were finalized from training and validation curves before accessing the formal test releases. Each release fixed the split, checkpoint hash, inference stride, batch size, and output location, and it could be consumed once. These controls separate development feedback from final evaluation and reduce the risk of test-set adaptation.

2.4. Geo-MSO Architecture

Geo-MSO is a U-Net-like encoder–decoder with base width 32, four downsampling stages, skip connections, and a 512-channel bottleneck. Local–spectral operator blocks are inserted only at 1/8 and 1/16 input resolution. At these scales, a depthwise 5×5 local branch represents building edges and shear layers, while a channel-wise spectral branch retains 6 and 12 low Fourier modes, respectively. A 1×1 projection mixes channels after the inverse transform. Group normalization, GELU activations, residual multilayer perceptrons, and FiLM conditioning complete each block.

For a feature tensor z, the spectral path applies a real two-dimensional Fourier transform, multiplies retained modes by learned complex channel-wise weights, transforms back to physical space, and performs pointwise channel mixing. This factorization is less expensive than learning dense input–output channel weights for every retained mode. It is

not mathematically equivalent to a full FNO layer; it is a memory-conscious global mixing path paired with explicit local convolution.

The conditioning vector $c = [\cos\theta, \sin\theta, u\tau/0.21]$ is projected to per-channel scale and bias. A conditioned normalized feature h is transformed as $(1 + \gamma(c)) \odot h + \beta(c)$. Thus, forcing modulates spatial features continuously without treating FLUX as a category. This describes the implemented mechanism, not an established causal benefit: the existing no-FiLM validation ablation does not isolate unseen-direction performance.

The decoder produces four means and four log Laplace scales for $t \in \{u, v, U, k\}$, combining shared-representation multi-task learning with an observation-scale estimate [23,27]. Log scales are clipped to $[-7, 5]$. Geo-MSO has 14,618,408 trainable parameters. Figure 3 summarizes the data path, and Table 3 documents differences among the compared architectures.

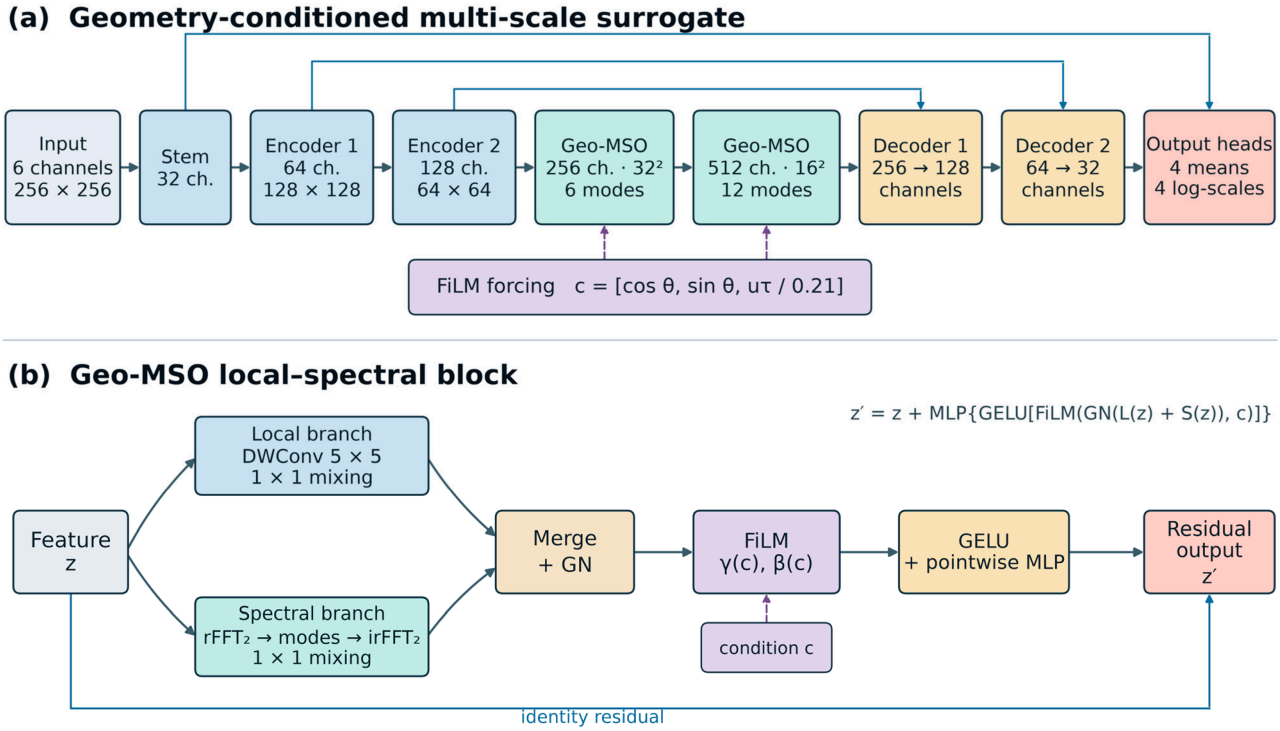


Figure 3. Geo-MSO architecture. Six scale-consistent input channels pass through a U-Net-like hierarchy. Local–spectral blocks operate at 1/8 and 1/16 resolution, receive FiLM forcing conditions, and feed four mean and four log-scale output heads. Dark solid arrows indicate the direction of feature propagation. Blue arrows represent skip connections in (**a**) and the identity residual connection in (**b**), while purple dashed arrows indicate FiLM conditioning. Different box colors distinguish the input features (light gray), encoding and local processing modules (light blue), Geo-MSO and spectral processing modules (cyan), decoding and feature fusion modules (light yellow), FiLM conditioning modules (purple), and output modules (pink).

Table 3. Neural architectures compared under a common data, training, and evaluation protocol.

Model	Spatial Mixing	Conditioning	Parameters
U-Net	Residual convolutional encoder–decoder	Input channels only	12.291 M
FNO	Four full-resolution dense spectral blocks; 16 modes/axis	Input channels only	9.489 M
Geo-MSO	Local + channel-wise spectral blocks at 1/8 and 1/16; 6/12 modes	Input channels + FiLM	14.618 M

2.5. Objective Function and Full-Field Reconstruction

For each target, the masked Laplace negative log-likelihood, omitting constants, is the mean of $\exp(-s)|\mu - y| + s$ over valid pixels. Four nondimensional target losses are

weighted equally. A first-order masked gradient loss compares horizontal and vertical finite differences of predicted and target means and is added with weight 0.1. Because both terms operate on nondimensional targets on the common 1 m model grid, the loss does not directly mix physical units.

The gradient term is a spatial regularizer, not a Navier–Stokes residual. Pedestrian-level slices omit vertical velocity and vertical gradients, so horizontal divergence cannot establish three-dimensional mass conservation. We therefore report the error between predicted and target horizontal divergence only as a structural diagnostic.

Training samples four random physical windows per case per epoch. Validation uses two deterministic low-discrepancy windows per case. Full fields are reconstructed with 256×256 windows, stride 128, and two-dimensional Hann overlap weights floored at 10^{-3} to avoid zero boundary denominators. The inference entry point reads geometry and forcing only; target NetCDF files are not opened during latency measurement.

2.6. Baselines, Optimization, and Checkpoint Selection

The U-Net baseline uses the same six channels, model grid, encoder–decoder depth, output heads, objective, sampling, and reconstruction protocol as Geo-MSO, but it excludes spectral blocks and FiLM. It has 12.291 million parameters. The FNO baseline uses width 48, four dense spectral blocks, and 16 modes per axis at full 256×256 resolution, followed by matched output heads; it has 9.489 million parameters. Geo-MSO has 14.618 million parameters and additional feature-level conditioning. The shared protocol controls several comparison factors but does not equalize capacity or conditioning; overall differences cannot be attributed solely to local–spectral mixing. A lightweight CNN, mean field, and nearest-neighbor retrieval were development checks, not three-seed formal-test competitors.

All formal neural models were trained with AdamW, initial learning rate 3×10^{-4}, weight decay 10^{-4}, cosine annealing to 10^{-6}, batch size 4, and gradient accumulation of two steps. Mixed-precision FP16 was used on one NVIDIA RTX 4090 D. Training could continue to 200 epochs and stopped after 30 validation epochs without improvement. The selected checkpoint minimized the validation macro-average composite loss. Random states for Python, NumPy, PyTorch, CUDA, and the data loader were recorded for seeds 20260904, 20260905, and 20260906. Checkpoints include the optimizer, scheduler, gradient scaler, early-stopping state, and random states, allowing continuity-checked recovery.

The measured server environment was Ubuntu 22.04, Python 3.10.21, PyTorch 2.5.1 + cu118, NumPy 2.0.2, SciPy 1.14.1, NetCDF4 1.7.2, cuDNN 9.1.0, one RTX 4090 D with 24,564 MiB memory, 16 visible CPU cores, and approximately 62 GiB memory. A full server audit read 538 cases and 3218 files with zero read failures for core NetCDF, terrain, and profile products.

2.7. Metrics and Statistical Reporting

All full-field metrics are first aggregated within each valid case and then macro-averaged across cases, preventing large domains from dominating solely through pixel count. Mean absolute error (MAE) measures typical magnitude error; root mean square error (RMSE) gives extra weight to localized large errors; and relative L2 normalizes the field error by target energy. Component MAEs describe directional biases, while vector MAE evaluates joint horizontal velocity error.

Direction error uses the periodic angular difference only where the target horizontal speed is at least 0.10 m s^{-1}. Near-building MAE is evaluated in fluid pixels with $0 < d \leq 8$ m. High-gradient MAE uses the highest 10% of target Uped gradient magnitude

within each case and is an evaluation-only diagnostic. Horizontal-divergence error is the MAE between the predicted and target two-dimensional divergence on the 1 m model grid.

For the Laplace heads, 90% interval coverage is compared with 0.90; higher is not automatically better. Spatial Spearman correlation between scale and absolute error assesses error ranking separately. Formal results are the mean ± sample SD across seeds 20260904, 20260905, and 20260906. Paired baseline-minus-Geo-MSO differences use the same seed and a descriptive two-sided Student-t 95% interval with two degrees of freedom. These intervals describe training variability on a fixed test set, not population uncertainty or high-powered significance tests. Reported percentage error reductions use the baseline denominator: 100 × (Ebaseline − EGeo)/Ebaseline. Detailed per-seed evaluation metrics and paired comparisons are provided in Supplementary Tables S1–S3 and S7–S8.

3. Results

3.1. Convergence and Training Cost

The 50-epoch screening run showed that Geo-MSO validation loss continued to improve by 8.92% over epochs 40–49, triggering the prespecified continuation rule. Its final best epochs were 49, 53, and 62; training ended at epochs 79, 83, and 92 after patience was exhausted. The corresponding U-Net and FNO checkpoints converged earlier on average. Across all fifteen formal trainings—three Geo-MSO splits and two geometry baselines, each with three seeds—the recorded epoch compute was 7.09 GPU h, excluding exploratory runs, preprocessing, evaluation, transfer, and ablation studies. Table 4 summarizes the training epochs, best-checkpoint epochs, and training times for the formal model comparisons.

Table 4. Formal training cost on one RTX 4090 D. Values are the mean ± sample SD over three seeds.

Model/Split	Epochs	Best Epoch	Time Per Seed	Three-Seed Total
Geo-MSO/geometry	85.7 ± 6.7	54.7 ± 6.7	42.78 ± 3.20 min	2.14 h
U-Net/geometry	76.3 ± 9.6	45.3 ± 9.6	33.54 ± 4.50 min	1.68 h
FNO/geometry	61.7 ± 10.1	30.7 ± 10.1	26.00 ± 4.25 min	1.30 h
Geo-MSO/city	52.3 ± 0.6	21.3 ± 0.6	15.58 ± 0.41 min	0.78 h
Geo-MSO/wind	72.7 ± 8.1	41.7 ± 8.1	23.94 ± 2.84 min	1.20 h

3.2. Primary Held-Out-Geometry Comparison

On the 81-case geometry_grouped test, Geo-MSO achieved Uped MAE of 0.127498 ± 0.000457 m s^{-1}, compared with 0.130305 ± 0.001613 m s^{-1} for U-Net and 0.136534 ± 0.000876 m s^{-1} for FNO. The baseline-relative reductions are 2.15% and 6.62%, respectively. Table 5 presents all eight primary diagnostics, including near-building and high-gradient errors previously described only in text. These are mixed-domain means over 34 Idealized and 47 Realistic cases; Table 6 separates the families for all three models.

Table 5. All eight primary held-out-geometry error diagnostics on the same 81 cases. Values are the case-macro mean ± sample SD across three seeds; lower is better. The combined test includes 34 Idealized and 47 Realistic cases.

Error Diagnostic/Unit	Geo-MSO	U-Net	FNO
Wind-speed MAE (m s^{-1})	0.12750 ± 0.00046	0.13030 ± 0.00161	0.13653 ± 0.00088
Wind-speed RMSE (m s^{-1})	0.16057 ± 0.00140	0.16343 ± 0.00147	0.17077 ± 0.00148
Wind-speed relative L2 (dimensionless)	0.35123 ± 0.00308	0.35750 ± 0.00403	0.37719 ± 0.00310
TKE MAE (m^2 s^{-2})	0.01855 ± 0.00021	0.01882 ± 0.00034	0.01965 ± 0.00005
Horizontal-vector MAE (m s^{-1})	0.22523 ± 0.00053	0.22826 ± 0.00253	0.23746 ± 0.00185
Near-building speed MAE (m s^{-1})	0.12494 ± 0.00031	0.12724 ± 0.00235	0.13418 ± 0.00066

Table 5. *Cont.*

Error Diagnostic/Unit	Geo-MSO	U-Net	FNO
High-gradient speed MAE (m s^{-1})	0.13393 ± 0.00104	0.13571 ± 0.00109	0.14340 ± 0.00100
Horizontal-divergence error (s^{-1})	0.01579 ± 0.00007	0.01643 ± 0.00016	0.01846 ± 0.00032

Speed and vector errors are in m s^{-1}, TKE MAE in m^2 s^{-2}, relative L2 is dimensionless, and horizontal-divergence error is in s^{-1}. Near-building and high-gradient regions follow Section 2.7. Full primary per-seed metrics and paired intervals are provided in Supplementary Tables S1–S3 and S7–S8.

Table 6. Same primary geometry test stratified by dataset family for all three models. Mean ± sample SD across the three training seeds; all velocity errors are in m s^{-1}. This is post hoc reporting of existing frozen predictions, not a new model-selection protocol.

Subset	Model	Speed MAE	Speed RMSE	Relative L2	Vector MAE
Idealized n = 34	Geo-MSO	0.03542 ± 0.00108	0.04606 ± 0.00128	0.13940 ± 0.00387	0.05115 ± 0.00125
Idealized n = 34	U-Net	0.03772 ± 0.00161	0.04916 ± 0.00211	0.14826 ± 0.00585	0.05453 ± 0.00228
Idealized n = 34	FNO	0.04982 ± 0.00211	0.06460 ± 0.00277	0.19869 ± 0.00790	0.07532 ± 0.00324
Realistic n = 47	Geo-MSO	0.19410 ± 0.00020	0.24340 ± 0.00155	0.50446 ± 0.00307	0.35115 ± 0.00146
Realistic n = 47	U-Net	0.19728 ± 0.00291	0.24609 ± 0.00180	0.50887 ± 0.00484	0.35394 ± 0.00565
Realistic n = 47	FNO	0.19926 ± 0.00018	0.24757 ± 0.00096	0.50632 ± 0.00193	0.35475 ± 0.00124

For FNO-minus-Geo-MSO, all eight primary error differences were positive in each of the three seeds, and their descriptive 95% intervals excluded zero. The Uped MAE difference was 0.009036 m s^{-1} with interval [0.006505, 0.011567]. This supports an aggregate difference between the complete implementations under the frozen protocol, not the causal effect of a single spatial-mixing component. Capacity, conditioning, and architecture differ, and the intervals do not estimate generalization uncertainty over new city populations.

The U-Net comparison was smaller. Geo-MSO had lower Uped MAE in all three seeds, but the paired mean difference of 0.002807 m s^{-1} had interval [−0.002263, 0.007877]. Most other U-Net intervals included zero; horizontal-divergence error had a positive interval, 0.000644 s^{-1} [0.000298, 0.000989]. We therefore report a modest aggregate mean reduction rather than broad statistical superiority. Figure 4 separates family-level results and explicitly labels the differing auxiliary test populations.

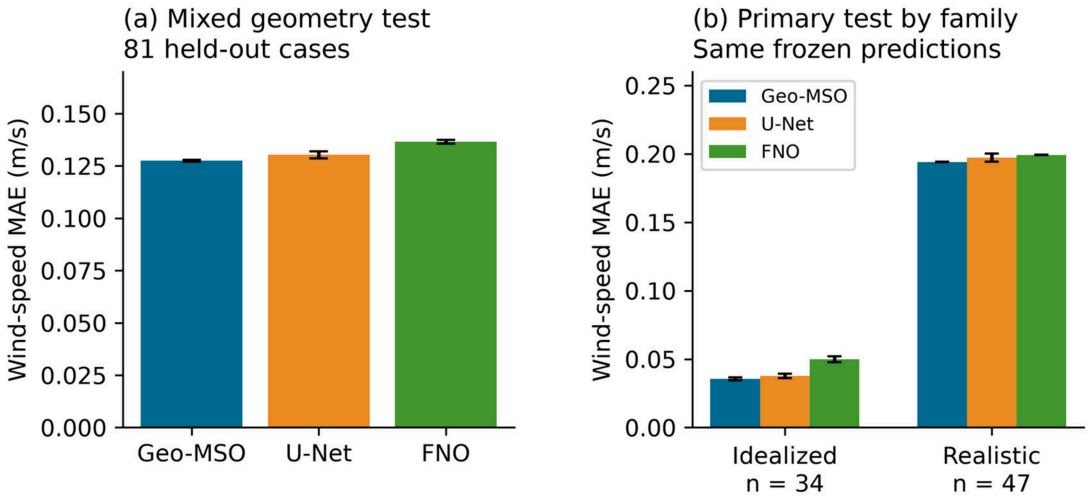


Figure 4. *Cont.*

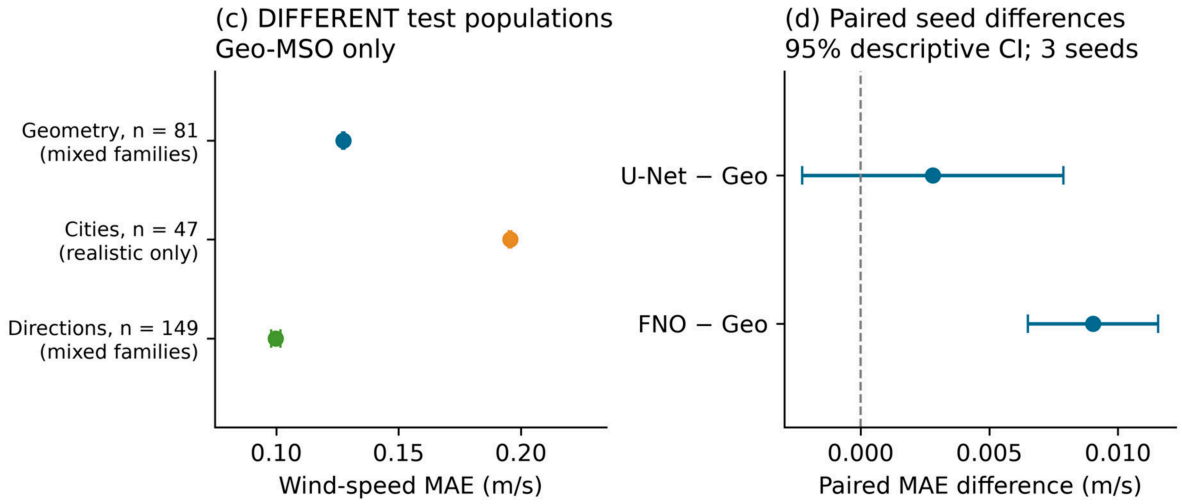


Figure 4. (**a**) Primary mixed-family geometry comparison. (**b**) The same test separated into 34 Idealized and 47 Realistic cases for all models. (**c**) Geo-MSO on three differently composed tests (81 geometries, 47 cities, 149 direction cases); these differently composed tests do not provide a controlled ranking of difficulty from their absolute errors. (**d**) Primary paired seed differences with descriptive Student-t 95% intervals. Error bars in (**a**–**c**) show sample SD across three seeds, not uncertainty over cities.The vertical dashed line in panel (**d**) indicates zero paired MAE difference.

3.3. Unseen Cities, Wind Directions, and Dataset Families

Geo-MSO obtained `Uped` MAEs of 0.195584 ± 0.000531 m s^{-1} on `city_grouped` and 0.099822 ± 0.001848 m s^{-1} on `wind_held_out`. The city result is numerically worse than the wind result, but the test sets differ in size, family, field extent, and composition; their absolute values do not isolate causal difficulty. In particular, 40 of the 47 city-test cases are from Melbourne, so the city aggregate is not a geographically balanced estimate. Table 7 presents the auxiliary generalization results for the geometry-, city-, and direction-held-out test populations.

Table 7. Geo-MSO auxiliary generalization results. Velocity errors are in m s^{-1}; nominal coverage is 0.90. These tests have different case populations and are not a controlled ranking of transfer difficulty.

Test	Cases	Uped MAE	RMSE	Rel. L2	Vector MAE	90% Coverage
Geometry held out	81	0.12750 ± 0.00046	0.16057 ± 0.00140	0.35123 ± 0.00308	0.22523 ± 0.00053	0.83774 ± 0.01710
Cities held out	47	0.19558 ± 0.00053	0.24287 ± 0.00192	0.51712 ± 0.00278	0.34484 ± 0.00405	0.93893 ± 0.01068
Directions held out	149	0.09982 ± 0.00185	0.12749 ± 0.00197	0.30422 ± 0.00524	0.17662 ± 0.00312	0.83365 ± 0.03139

Family stratification uses the existing primary-test predictions without new tuning. The 34 Idealized cases gave Geo-MSO a mean Uped MAE of 0.035423 m s^{-1} versus 0.037723 for U-Net and 0.049820 for FNO; the corresponding 47 Realistic means were 0.194104, 0.197278, and 0.199263 m s^{-1} (Table 6). Geo-MSO beat FNO in 33/34 Idealized cases (97.1%) but only 24/47 Realistic cases (51.1%), using each case's mean across seeds. For Realistic cases, mean FNO-minus-Geo-MSO was 0.005158 m s^{-1} but median was 0.000099 m s^{-1}. The mean improvement is therefore not a consistently positive advantage across realistic neighborhoods. These descriptive case fractions are not independent-city significance tests.

The validation case CA-Sunset was selected by the frozen realistic-maximum-error rule, not by the formal-test results. The Geo-MSO, U-Net, and FNO Uped MAEs were 0.5678, 0.5055, and 0.7374 m s^{-1}. Figure 5 enlarges the target and three predictions with shared color limits; all smooth narrow high-speed street-canyon jets, and U-Net is better in this case. A low predicted wind speed in a narrow passage must therefore not be read as

the absence of a high-wind hazard. Whole-field MAE can dilute a localized peak error, and the predicted Laplace scales cannot reliably identify such failures (Section 3.4).

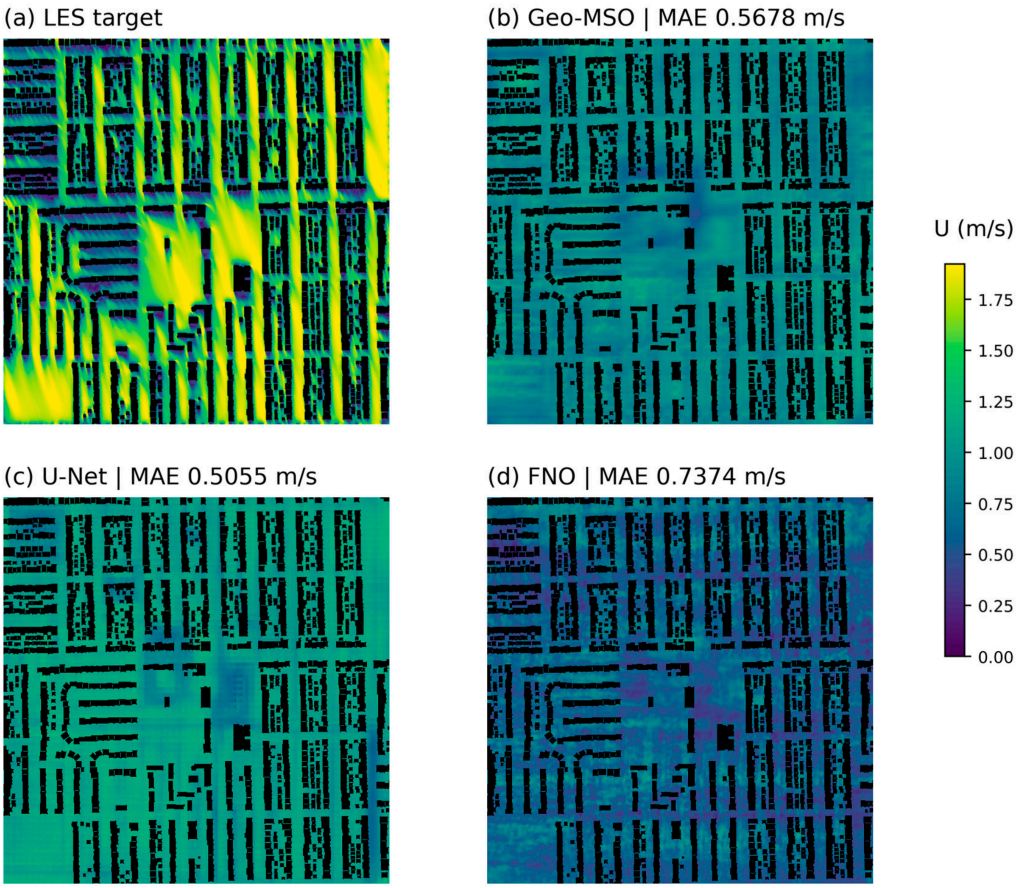


Figure 5. CA-Sunset validation failure, selected by the frozen realistic-maximum-error rule. (**a**) LES target Uped; (**b**–**d**) Geo-MSO, U-Net, and FNO predictions. The four enlarged fields share identical color limits; black regions are masked. The saved rendered predictions have only been reassembled for legibility, with no change to prediction values or case selection. Narrow high-speed corridors visible in the target are smoothed by all models.

Exploratory correlations between error and 17 input-only geometry descriptors showed that field size and dataset family explained much of the raw mixed-domain association. After controlling for log output-pixel count and family, most relationships weakened markedly. Within realistic wind cases, mean height, 16/64 m texture measures, and windward area retained moderate associations; these are mechanism hypotheses for future data collection, not independent causal explanations.

3.4. Ablations, Transfer Learning, and Uncertainty

Table 8 exposes the available single-factor validation ablations (81 cases, seed 20260904), with full metrics in Table S5. Removing signed distance increased Uped and TKE MAE but slightly reduced vector error, so its benefit is metric-specific. A single operator stage reduced scalar MAE but worsened TKE, vector, and divergence errors. Removing FiLM also reduced scalar MAE in this seed (0.129239 versus 0.130524 m s^{-1}) while worsening vector error and changing 90% coverage (0.817871 versus 0.861417). This does not support a general point-accuracy benefit from FiLM. The uncertainty-head ablation showed several modest error trade-offs, not calibrated operational risk prediction. These studies are single-seed, validation-only, and not parameter-matched; no matched local-only

versus spectral-only experiment was performed. The roles of individual mechanisms and unseen-direction FiLM benefits remain unresolved.

Table 8. Available single-factor geometry-validation ablations (81 cases; seed 20260904). These are not formal tests, parameter-matched studies, or reliable rankings of individual component value. Full metrics and the three-seed gradient study are supplied in Tables S5 and S6.

Variant	Speed MAE ($m\ s^{-1}$)	TKE MAE ($m^2\ s^{-2}$)	Vector MAE ($m\ s^{-1}$)	Div. Error (s^{-1})	90% Coverage
Full model	0.13052	0.01968	0.23928	0.01713	0.86142
No signed distance	0.13189	0.02011	0.23925	0.01721	0.84983
One operator stage	0.12903	0.02014	0.24312	0.01745	0.84955
No FiLM	0.12924	0.01993	0.24021	0.01681	0.81787
No gradient term	0.12773	0.01970	0.24372	0.01758	0.85855
No scale heads	0.13073	0.01998	0.24424	0.01763	Not defined

The hypothesis that the gradient term would improve scalar high-gradient recovery was not supported by the three-seed validation ablation. Mean high-gradient Uped MAE was 0.143130 m s^{-1} without the term versus 0.143607 with it. Nevertheless, vector, u-component, and horizontal-divergence errors were lower with the term in all three paired seeds (Table S6). We retain it as a multi-output spatial regularizer and do not claim it recovers narrow jets or enforces three-dimensional mass conservation.

The Idealized-to-Realistic transfer study used 176/24/24 source cases and a nested 25% realistic training subset of 66 cases, with 47 fixed city-grouped target validation cases. Zero-shot `Uped` MAE was 0.241245 m s^{-1}. Fine-tuning improved substantially over zero-shot use, but at each tested learning rate, it did not beat a same-budget random initialization. At learning rate 3×10^{-4}, pretrained and scratch MAEs were 0.199258 and 0.197971 m s^{-1}, respectively, a 0.650% disadvantage for pretraining. Pretraining reached its best checkpoint at epoch 5 instead of 30 and used 50.8% less target-domain epoch compute, although including 1207.02 s of source pretraining removed a single-task end-to-end compute advantage. Training FiLM plus heads only and using discriminative learning rates also did not beat scratch. These are negative results for the tested naïve transfer strategies, not for domain adaptation in general. Table 9 summarizes the validation-set results and training costs of the evaluated transfer strategies, while Figure 6 illustrates their target-domain convergence curves.

Table 9. Idealized-to-Realistic transfer on the fixed target validation set using the nested 25% Realistic training subset. These are development results, not new formal-test comparisons.

Strategy	Best Epoch	Target Time	Uped MAE	TKE MAE	Vector MAE
Zero-shot	—	0.00 min	0.241245	0.038518	0.452642
Pretrained full FT, 1×10^{-4}	10	4.17 min	0.202863	0.026662	0.382228
Scratch, 1×10^{-4}	41	8.25 min	0.201231	0.026691	0.390447
Pretrained full FT, 3×10^{-4}	5	3.74 min	0.199258	0.027018	0.379410
Scratch, 3×10^{-4}	30	7.60 min	0.197971	0.027201	0.381936
FiLM + heads only	99	12.50 min	0.207067	0.029976	0.391897
Discriminative LR	31	6.99 min	0.203364	0.026700	0.380844

Velocity errors are in m s^{-1} and TKE MAE in $m^2\ s^{-2}$. FT: full fine-tuning; LR: learning rate.

Nominal 90% interval coverage was 0.837744, 0.938928, and 0.833651 on the geometry, city, and wind tests, respectively. The heads therefore under-covered geometry and wind while over-covering the city test. In CA-Sunset, predicted scale had near-zero rank correla-

tion with actual pixel error. The heteroscedastic outputs may regularize point prediction, but they are not sufficiently calibrated or spatially informative for risk localization.

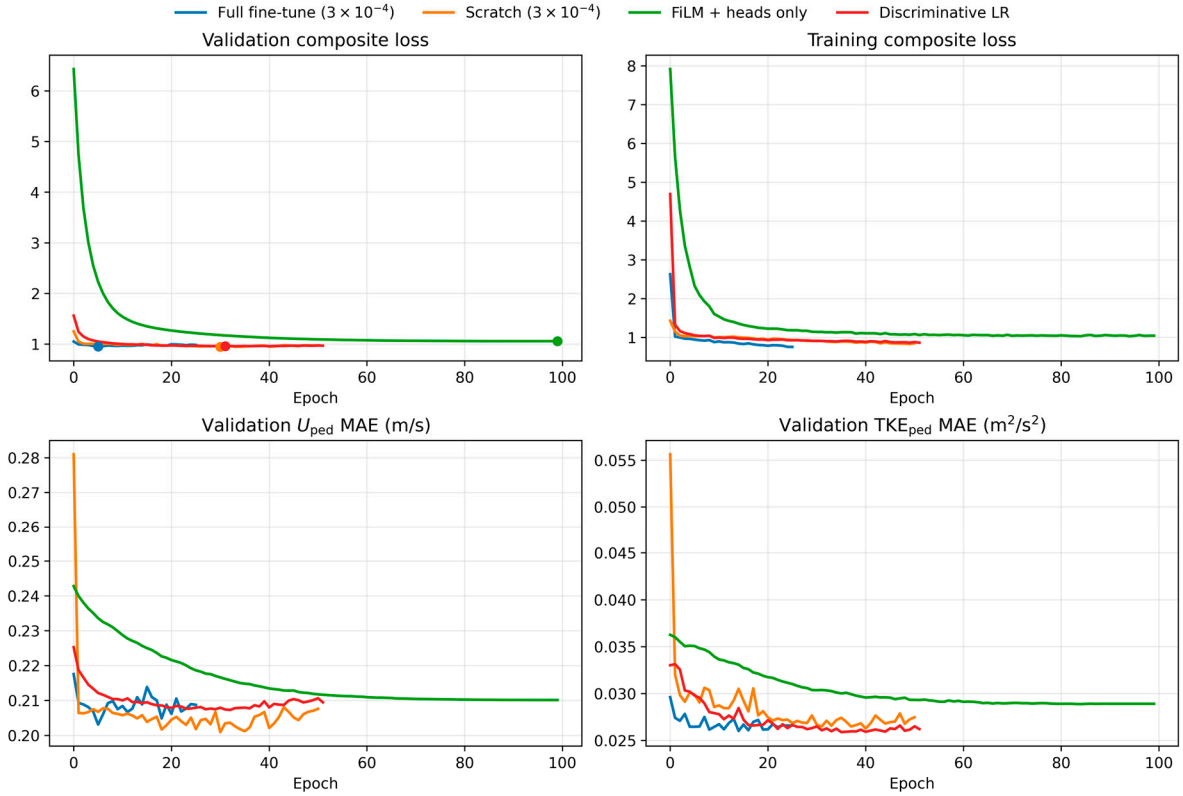


Figure 6. Idealized-to-Realistic transfer remediation curves on the fixed Realistic validation set. Pretraining reaches the selected target checkpoint earlier, but none of the tested transfer strategies improves final Uped MAE over the matched scratch model.

3.5. Deployment Latency

Latency was measured after model selection on an otherwise idle RTX 4090 D using PyTorch 2.5.1 + cu118 and FP16. Patch-forward timings used 20 warm-up runs and 100 measured repetitions with CUDA synchronization. Batch-one and batch-four medians were 8.095 and 8.253 ms, respectively, for 256 × 256 patches. Full-field cases used one warm-up and five measured repetitions and included file access, geometry preprocessing, transfers, inference, and overlap stitching. Table 10 reports end-to-end latency, tile count, GPU time, preprocessing time, and peak memory for representative full-field predictions.

Table 10. Warm end-to-end full-field latency on an RTX 4090 D with FP16. Input preparation, transfer, inference, and stitching are included; no matched CFD speedup is inferred.

Case	Field	Tiles	Median/P90	GPU	Preprocess	Peak Memory
VS64-CL-28_d00	60 × 120	1	113.69/116.04 ms	9.13 ms	97.38 ms	158.15 MiB
UA16_d00	120 × 200	2	209.23/212.22 ms	9.37 ms	185.21 ms	220.53 MiB
US-Det-U6_d90	240 × 384	6	160.18/165.03 ms	15.70 ms	117.42 ms	340.78 MiB
TA-DC-U1_d00	576 × 624	25	886.43/906.38 ms	63.99 ms	677.88 ms	346.78 MiB
US-WestPhoenix	1008 × 1008	64	1747.47/1784.24 ms	123.79 ms	1346.86 ms	346.90 MiB

End-to-end median time ranged from 113.69 ms for a one-tile 60 × 120 field to 1747.47 ms for a 1008 × 1008 field reconstructed from 64 tiles. For the latter, GPU forward time was 123.79 ms, whereas preprocessing required 1346.86 ms. The dominant engineering opportunity is therefore CPU geometry preparation and caching rather than additional GPU

kernel optimization. No speedup factor relative to CFD is reported because no matched CFD timing was collected for the same case, grid, hardware, and convergence criterion.

4. Discussion

4.1. What the Evidence Supports

The contribution is a combined, auditable benchmark and task-specific model, not a new physical theory or proof of the isolated value of hybrid spatial mixing. Our aggregate primary results favor Geo-MSO over the implemented FNO, but the networks differ in capacity and conditioning, and the available ablations do not establish causal attribution. U-FNO already demonstrates local–spectral hybridization in another flow task [17], while FNO, DeepONet, and GINO supply broader operator-learning formulations [14,16,18]. Geo-MSO applies a factorized coarse-resolution mixing path to fixed-grid pedestrian statistics; it does not demonstrate mesh-independent operator convergence, irregular-grid inference, or three-dimensional PDE solution.

Relative to urban-wind studies, the distinction is primarily evaluation scope and data semantics. Height-raster CNN screening and accuracy–cost trade-offs are established by Clemente et al. and FastFlow [7,8]; morphology/direction prediction, city generalization, hierarchical refinement, and seamless reconstruction are addressed by Lu et al., Chen et al., Snaiki et al., and Huang et al. [10–13]. Our study puts grouped layout identities, Val/base pairs, physical scale, family-specific results, negative transfer outcomes, and seed-level paired tables into one reproducible protocol. Gao et al. use sparse flow sensors [20], so their reconstruction task is not a geometry-only comparison. Different datasets, fields, splits, and timing boundaries preclude numerical cross-paper rankings. Within our common protocol, U-Net remains a strong baseline: most paired intervals overlap zero, and it is better on the displayed CA-Sunset failure.

The scientific implication is that aggregate accuracy is conditional on morphology and evaluation design. The primary family table and near-zero Realistic median reveal what the mixed-domain average obscures: the method is not consistently better across complex neighborhoods. The fixed physical window, target nondimensionalization, and grouping of shared layouts make this conclusion interpretable and reproducible. FLUX, Val-*, and case-map handling are reusable processing safeguards for UrbanTALES [25], rather than evidence of universal superiority over other published models. Model/data co-design must be assessed using these controls together with local errors, not a scalar leaderboard alone.

4.2. Realistic-Neighborhood Bottleneck

The idealized–realistic error gap and CA-Sunset failure indicate that the remaining problem is concentrated around realistic morphology and narrow jets. Coarse spectral mixing can convey long-range context, but the common 1 m output grid and data-weighted regression objective still favor spatially smoother predictions. Rare peak structures occupy few pixels, may be sensitive to geometry rasterization, and are insufficiently represented by global average objectives.

The observed failures motivate hypotheses rather than tested remedies. Geometry-adaptive multiresolution decoding and input-only canyon/corner sampling could target poorly represented peaks. Domain-balanced batches or adaptation should be compared with strong scratch training, because naïve pretraining did not improve final target accuracy. Physical residual learning [21,22] would need additional state variables and boundary information; simply minimizing horizontal divergence on a slice is not sufficient. Held-out realistic calibration or ensembles could complement heteroscedastic heads [23,24],

but exchangeability and domain shift must be checked before any coverage guarantee is invoked.

4.3. Implications for Building Layout and Urban-Wind Screening

The implementation-ready result is a reproducible method for producing pedestrian-height mean-field estimates from compatible geometry and forcing, with measured inference latency. In early-stage planning, these estimates can help visualize broad ventilation pathways and identify candidate layouts or forcing cases for subsequent examination [5,8]. Proposed uses such as rapid variant filtering remain conditional on input compatibility and local validation; this study did not validate a design optimizer, a constructed intervention, or an operational decision rule.

A cautious workflow is to verify geometric units, coordinate orientation, forcing convention, and morphology coverage; inspect speed, vector, and near-building maps together; and flag narrow passages, corners, and sensitive occupied locations for higher-fidelity analysis irrespective of a low predicted value. In such locations, do not use the surrogate minimum or its uncalibrated interval as a safety certificate. Changes selected by the surrogate should be re-evaluated using validated CFD or wind-tunnel/field evidence before final design. Site-specific wind climate, directional probabilities, gust factors, comfort/safety criteria, thermal conditions, and applicable regulations are outside the present dataset and must be evaluated separately.

Measured full-field latency supports interactive or batch screening, but it does not establish a CFD speedup ratio. Geometry preprocessing dominates the largest measured case, suggesting input caching as an implementation improvement. No validated intervention magnitude, canyon-width threshold, or engineering safety margin can be prescribed from the current results. This restricted role aligns the contribution with building-design exploration while preserving established CFD verification practice [1,2].

4.4. Limitations

First, the target is a two-dimensional pedestrian-height LES statistic. Vertical velocity, pressure, and vertical gradients are unavailable in the local release, so the model cannot establish three-dimensional mass or momentum consistency. Second, the number of independent geometries is smaller than the number of cases because directions and resolution pairs share sites. Grouped splits mitigate leakage but also increase variance.

Third, the city test is geographically imbalanced, with 40/47 cases from Melbourne, and auxiliary tests lack a full three-model matched comparison. They are not population-weighted estimates across 73 cities. Fourth, most component ablations are single-seed validation studies; unequal parameter counts and conditioning remain comparison confounders, and isolated local/spectral and capacity-matched tests are absent. Fifth, the Laplace heads are not calibrated for operational risk decisions or reliable peak-error localization. Sixth, timing is hardware- and implementation-specific; absent matched CFD wall-clock data, an acceleration ratio is not defensible.

Finally, all training data originate from one simulation system and its modeling assumptions. Agreement with UrbanTALES LES should not be equated with validation against wind-tunnel or field measurements. A deployment study should examine numerical-model shift, measurement noise, atmospheric-stability regimes, and geometry acquisition errors.

5. Conclusions

The following conclusions are separated into empirical scientific findings, applied scope, and future research; they do not claim a new urban-flow law.

1. Scientific finding—combined pipeline. A physical-scale-consistent representation and grouped frozen protocol enabled reproducible comparison across 538 UrbanTALES cases. On the 81-case mixed geometry test, Geo-MSO reduced mean Uped MAE by 6.62% relative to the FNO baseline and 2.15% relative to U-Net, using each baseline as the percentage denominator. All eight primary FNO differences favored Geo-MSO across three seeds; most U-Net intervals included zero. These are complete-model comparisons, not isolated component effects.
2. Scientific finding—conditional generalization. The mixed-test mean does not establish consistent Realistic-domain superiority: Geo-MSO beat FNO in 24/47 realistic cases (51.1%), with a median difference of only 0.000099 m s^{-1}. Narrow street-canyon jets, the imbalanced unseen-city test, and uncertainty calibration remained limitations. Idealized pretraining accelerated target convergence without improving final Realistic accuracy over matched scratch training. The evidence supports morphology-qualified evaluation and reporting of negative results rather than unconditional model dominance.
3. Applied finding. End-to-end inference took approximately 0.114–1.747 s for the measured 1- to 64-tile fields on an RTX 4090 D. The implemented pipeline is suitable for preliminary mean-field screening within compatible data conditions, but low predictions in narrow passages cannot exclude high-wind exposure. Final comfort/safety decisions and candidate building interventions require independently validated analysis and site wind-climate information; no matched CFD speedup or certified design recommendation is claimed.
4. Future research. A newly reserved, geographically balanced Realistic test set, capacity- and conditioning-controlled comparisons, geometry-adaptive peak recovery, and independent numerical/measurement validation are needed. Future uncertainty calibration and transfer experiments should be evaluated on untouched realistic cases and reported alongside failure-localization and full-pipeline cost metrics.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/buildings16214233/s1: Table S1: Geo-MSO: all primary-test per-seed diagnostics (81 cases); Table S2: U-Net: all primary-test per-seed diagnostics (81 cases); Table S3: FNO: all primary-test per-seed diagnostics (81 cases); Table S4: Family-specific primary-test diagnostics for all three models; Table S5: Existing single-seed validation ablations (81 geometry-validation cases); Table S6: Three-seed gradient ablation on the fixed validation set; Table S7: Primary paired comparisons: U-Net minus Geo-MSO; Table S8: Primary paired comparisons: FNO minus Geo-MSO. The supplementary material provides detailed per-seed evaluation results, family-stratified performance summaries, validation ablations, and paired model comparisons. The source code and reproducibility tests are available at https://github.com/Zyzrepositories/Geo_MSO_UrbanTALES_Buildings.

Author Contributions: Conceptualization, Y.Z. and Y.Y.; methodology, Y.Z.; software, Y.Z.; formal analysis, Y.Z.; investigation, Y.Z.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z. and Y.Y.; visualization, Y.Z.; supervision, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original UrbanTALES simulation data are publicly available from the official project portal at https://urbantales.vercel.app/ (accessed on 4 May 2026) and are described in the dataset publication [25]. The processing, training, and evaluation code developed for this study

is openly available in the authors' GitHub repository at https://github.com/Zyzrepositories/Geo_MSO_UrbanTALES_Buildings. Additional experimental results are provided in the accompanying Supplementary_Materials.docx.

Acknowledgments: The authors gratefully acknowledge the UrbanTALES project for making the urban turbulent airflow dataset publicly available. The authors also thank all individuals and institutions that contributed to the preparation of this study. All authors reviewed and approved the final manuscript and take full responsibility for its content.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

Abbreviation	Definition
CFD	Computational fluid dynamics
FNO	Fourier neural operator
Geo-MSO	Geometry-conditioned multi-scale operator
LES	Large-eddy simulation
MAE	Mean absolute error
RMSE	Root mean square error
SDF	Signed distance field
TKE	Turbulent kinetic energy

References

1. Tominaga, Y.; Mochida, A.; Yoshie, R.; Kataoka, H.; Nozu, T.; Yoshikawa, M.; Shirasawa, T. AIJ Guidelines for Practical Applications of CFD to Pedestrian Wind Environment around Buildings. *J. Wind Eng. Ind. Aerodyn.* **2008**, *96*, 1749–1761. [CrossRef]
2. Blocken, B. Computational Fluid Dynamics for Urban Physics: Importance, Scales, Possibilities, Limitations and Ten Tips and Tricks towards Accurate and Reliable Simulations. *Build. Environ.* **2015**, *91*, 219–245. [CrossRef]
3. Toparlar, Y.; Blocken, B.; Maiheu, B.; van Heijst, G.J.F. A Review on the CFD Analysis of Urban Microclimate. *Renew. Sustain. Energy Rev.* **2017**, *80*, 1613–1640. [CrossRef]
4. Calzolari, G.; Liu, W. Deep Learning to Replace, Improve, or Aid CFD Analysis in Built Environment Applications: A Review. *Build. Environ.* **2021**, *206*, 108315. [CrossRef]
5. Wu, Y.; Quan, S.J. A Review of Surrogate-Assisted Design Optimization for Improving Urban Wind Environment. *Build. Environ.* **2024**, *253*, 111157. [CrossRef]
6. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241. [CrossRef]
7. Low, S.J.; Raghavan, V.S.G.; Gopalan, H.; Wong, J.C.; Yeoh, J.; Ooi, C.C. FastFlow: AI for Fast Urban Wind Velocity Prediction. *arXiv* **2022**, arXiv:2211.12035. [CrossRef]
8. Clemente, A.V.; Giljarhus, K.E.T.; Oggiano, L.; Ruocco, M. Rapid Pedestrian-Level Wind Field Prediction for Early-Stage Design Using Pareto-Optimized Convolutional Neural Networks. *Comput.-Aided Civ. Infrastruct. Eng.* **2024**, *39*, 2826–2839. [CrossRef]
9. Wang, H.; Ma, W.; Niu, J.; You, R. Evaluating a Deep Learning-Based Surrogate Model for Predicting Wind Distribution in Urban Microclimate Design. *Build. Environ.* **2025**, *269*, 112426. [CrossRef]
10. Lu, J.; Li, W.; Hobeichi, S.; Azad, S.A.; Nazarian, N. Machine Learning Predicts Pedestrian Wind Flow from Urban Morphology and Prevailing Wind Direction. *Environ. Res. Lett.* **2025**, *20*, 054006. [CrossRef]
11. Chen, C.; Tian, G.; Qin, S.; Yang, S.; Geng, D.; Zhan, D.; Yang, J.; Vidal, D.; Wang, L.L. Generalization of Urban Wind Environment Using Fourier Neural Operator across Different Wind Directions and Cities. *arXiv* **2025**, arXiv:2501.05499. [CrossRef]
12. Snaiki, R.; Lu, J.; Li, S.; Nazarian, N. A Hierarchical Deep Learning Model for Predicting Pedestrian-Level Urban Winds. *Build. Environ.* **2026**, *294*, 114354. [CrossRef]
13. Huang, J.; Heaney, C.E.; Li, T.; Li, X.; Hughes, G.O.; van Reeuwijk, M. Inpainting U-Net for Seamless Pedestrian-Level Wind Prediction across Urban Morphologies. *arXiv* **2026**, arXiv:2607.02560. [CrossRef]
14. Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhattacharya, K.; Stuart, A.; Anandkumar, A. Fourier Neural Operator for Parametric Partial Differential Equations. In Proceedings of the 9th International Conference on Learning Representations, Virtual Conference, 3–7 May 2021.

15. Kovachki, N.; Li, Z.; Liu, B.; Azizzadenesheli, K.; Bhattacharya, K.; Stuart, A.; Anandkumar, A. Neural Operator: Learning Maps between Function Spaces with Applications to PDEs. *J. Mach. Learn. Res.* **2023**, *24*, 89.
16. Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; Karniadakis, G.E. Learning Nonlinear Operators via DeepONet Based on the Universal Approximation Theorem of Operators. *Nat. Mach. Intell.* **2021**, *3*, 218–229. [CrossRef]
17. Wen, G.; Li, Z.; Azizzadenesheli, K.; Anandkumar, A.; Benson, S.M. U-FNO—An Enhanced Fourier Neural Operator-Based Deep-Learning Model for Multiphase Flow. *Adv. Water Resour.* **2022**, *163*, 104180. [CrossRef]
18. Li, Z.; Kovachki, N.; Choy, C.; Li, B.; Kossaifi, J.; Otta, S.P.; Nabian, M.A.; Stadler, M.; Hundt, C.; Azizzadenesheli, K. Geometry-Informed Neural Operator for Large-Scale 3D PDEs. In *Advances in Neural Information Processing Systems 36*; NeurIPS: San Diego, CA, USA, 2023; Available online: https://arxiv.org/abs/2309.00583 (accessed on 4 May 2026).
19. Perez, E.; Strub, F.; de Vries, H.; Dumoulin, V.; Courville, A. FiLM: Visual Reasoning with a General Conditioning Layer. In *AAAI'18: AAAI Conference on Artificial Intelligence*; AAAI Press: Palo Alto, CA, USA, 2018; Volume 32. [CrossRef]
20. Gao, H.; Hu, G.; Zhang, D.; Jiang, W.; Tse, K.T.; Kwok, K.C.S.; Kareem, A. Urban Wind Field Prediction Based on Sparse Sensors and Physics-Informed Graph-Assisted Auto-Encoder. *Comput.-Aided Civ. Infrastruct. Eng.* **2024**, *39*, 1409–1430. [CrossRef]
21. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-Informed Machine Learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [CrossRef]
22. Li, Z.; Zheng, H.; Kovachki, N.; Jin, D.; Chen, H.; Liu, B.; Azizzadenesheli, K.; Anandkumar, A. Physics-Informed Neural Operator for Learning Partial Differential Equations. *ACM/IMS J. Data Sci.* **2024**, *1*, 9. [CrossRef]
23. Kendall, A.; Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In *Advances in Neural Information Processing Systems 30*; Curran Associates: Red Hook, NY, USA, 2017.
24. Angelopoulos, A.N.; Bates, S. Conformal Prediction: A Gentle Introduction. *Found. Trends Mach. Learn.* **2023**, *16*, 494–591. [CrossRef]
25. Nazarian, N.; Lu, J.; Lipson, M.J.; Hart, M.A.; Liu, S.; Krayenhoff, E.S.; Blunn, L.; Martilli, A. UrbanTALES: A Large-Eddy Simulation Dataset for Urban Canopy Layer Turbulence and Parameterization. *Bull. Am. Meteorol. Soc.* **2025**, *106*, E2461–E2478. [CrossRef]
26. Maronga, B.; Banzhaf, S.; Burmeister, C.; Esch, T.; Forkel, R.; Fröhlich, D.; Fuka, V.; Gehrke, K.F.; Geletič, J.; Giersch, S.; et al. Overview of the PALM Model System 6.0. *Geosci. Model Dev.* **2020**, *13*, 1335–1372. [CrossRef]
27. Caruana, R. Multitask Learning. *Mach. Learn.* **1997**, *28*, 41–75. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.