URBANGRAPH: PHYSICS-INFORMED SPATIO-TEMPORAL DYNAMIC HETEROGENEOUS GRAPHS FOR URBAN MICROCLIMATE PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

With rapid urbanization, predicting urban microclimates has become critical, as it affects building energy demand and public health risks. However, existing generative and homogeneous graph approaches fall short in capturing physical consistency, spatial dependencies, and temporal variability. To address this, we introduce UrbanGraph, a physics-informed framework integrating heterogeneous and dynamic spatio-temporal graphs. It encodes key physical processes—vegetation evapotranspiration, shading, and convective diffusion—while modeling complex spatial dependencies among diverse urban entities and their temporal evolution. We evaluate UrbanGraph on UMC4/12, a physics-based simulation dataset covering diverse urban configurations and climates. Results show that UrbanGraph improves R^2 by up to 10.8% and reduces FLOPs by 17.0% over all baselines, with heterogeneous and dynamic graphs contributing 3.5% and 7.1% gains. Our dataset provides the first high-resolution benchmark for spatio-temporal microclimate modeling, and our method extends to broader urban heterogeneous dynamic computing tasks.

1 INTRODUCTION

Urban microclimate prediction is crucial for urban sustainability and public health (Grant et al., 2025; He et al., 2024). This task represents a broad class of spatio-temporal urban physical field prediction problems, such as urban wind field simulation and pollutant dispersion forecasting. The core challenge of these problems is that the physical state at any point in urban space is determined by the collective interactions among numerous and diverse urban entities (e.g., buildings, vegetation) through time-varying physical processes such as radiation and convection (Coutts et al., 2013; de Abreu-Harbich et al., 2015; Irmak et al., 2017; Abd Elraouf et al., 2022). While high-fidelity physics-based numerical simulations, such as Computational Fluid Dynamics (CFD), are the standard approach for solving such problems, their immense computational overhead makes them infeasible for large-scale, time-series prediction tasks. Therefore, exploring computationally efficient data-driven methods to strike a balance between prediction accuracy and efficiency has become an essential research direction.

Although data-driven methods are promising, they still face challenges in accurately modeling the underlying physical processes. Station-based methods can only predict time series at discrete locations, neglecting spatial relationships and failing to generate continuous physical fields. Grid-based generative models are constrained by local receptive fields, which makes it difficult for them to capture long-range spatial dependencies (Carter et al., 2016; Kemppinen et al., 2024). Graph Neural Networks (GNNs) offer a more natural framework for modeling the spatial dependencies among urban entities. However, existing GNN-based approaches often lack physical consistency. They typically employ a uniform message-passing mechanism that cannot distinguish between different physical processes, such as vegetation evapotranspiration and building shading (Zhao et al., 2021). Furthermore, these methods struggle to model temporal variability. They mostly rely on a fixed graph structure, which is incapable of representing how physical processes evolve in real-time in response to changing environmental conditions. Consequently, there is a pressing need in the field for a unified framework capable of explicitly modeling multiple physical processes and their temporal evolution.

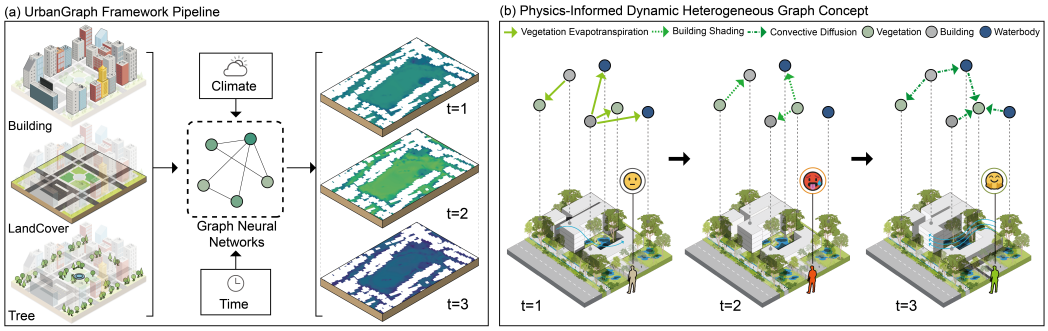

Figure 1: The UrbanGraph framework. (a) Overall pipeline: Geospatial data is converted into a graph structure, which is processed by a spatio-temporal GNN with climate and time characteristics to generate high-resolution predictions. (b) The physics-informed dynamic graph concept: Nodes represent urban entities, while edges—representing physical interactions like building shading and convective diffusion—are dynamically reconfigured over time to reflect changing environmental conditions.

Addressing this gap requires solving the following two core and orthogonal technical challenges. (i) A method must be designed to encode multiple, independent, and time-varying physical processes into a dynamic, heterogeneous graph structure. The fundamental difficulty lies in abstracting continuous physical fields (such as radiative transfer and fluid dynamics) into a discrete and computationally efficient graph topology, without losing critical physical information. (ii) A neural network architecture must be designed that can effectively process this complex sequence of graphs. This is highly challenging because the model must not only handle the dual dynamics of both node features and graph topology but also remain sensitive to the diverse types of relationships, each driven by different physical processes. Ultimately, the entire framework must strike a balance between physical interpretability and computational efficiency, while avoiding oversimplification of the urban morphology (Heo et al.).

To address these challenges, we propose the UrbanGraph framework,, illustrated in Figure 1. It first employs a physics-informed graph representation method to explicitly map multiple physical processes, such as shading and convection, into different types of dynamic edges, thereby transforming physical laws into a computable graph structure. Subsequently, we design a spatio-temporal graph network centered on a heterogeneous message-passing mechanism. This mechanism can assign dedicated learning functions to different physical processes (Schlichtkrull et al., 2017) and capture their dynamic evolution. The effectiveness of this approach is validated on a custom high-fidelity, physics-based simulation benchmark dataset. The results on prediction accuracy and computational efficiency across multiple targets suggest that this framework can be generalized to other urban computing tasks.

To summarize, we make the following contributions:

- We propose a physics-informed graph representation that explicitly encodes time-varying physical processes and static spatial relationships into the topology of a dynamic heterogeneous graph. This method offers a novel pathway for injecting temporally evolving domain knowledge into graph learning.
- We develop a dynamic heterogeneous graph neural network architecture for the UrbanGraph framework, which efficiently learns from the complex graph sequences we propose through a heterogeneous message-passing mechanism. Comprehensive experiments demonstrate that the architecture achieves state-of-the-art performance in both prediction accuracy and computational efficiency. Compared to four categories of baselines, it improves accuracy by up to 10.8% (in R^2) and efficiency by 17.0% (in FLOPs).
- We provide a new, well-validated perspective for modeling urban systems. Results quantitatively demonstrate that the heterogeneous and dynamic mechanisms are key to the performance improvement, contributing gains of 3.5% and 7.1%, respectively. Furthermore, the architecture's results on multiple target variables show strong generalization capabilities.

2 Related Work

Classical Microclimate Prediction Methods. Classical methods for urban microclimate prediction can be broadly categorized into two types. The first consists of physics-based simulation models, such as CFD and ENVI-met (Toparlar et al., 2017; Tsoka et al., 2018; Liu et al., 2021; Barros Moreira de Carvalho & Bueno da Silva, 2024). These methods offer high physical fidelity but suffer from immense computational overhead, making them impractical for large-scale, long-term time-series prediction tasks. The second category comprises data-driven approaches, including traditional machine learning (Arulmozhi et al., 2021; Alaoui et al., 2023) and grid-based deep learning models like CNNs (Kumar et al., 2021; Kastner & Dogan, 2023; Fujiwara et al., 2024). While these methods are computationally efficient, the former struggle to capture complex spatial dependencies, and the latter are constrained by the Euclidean data assumption, making them unable to process the inherently non-structural geometry of urban environments. To address these limitations, GNNs provide a more suitable framework, as their message-passing mechanism can directly model complex spatial dependencies (Kipf & Welling, 2016; Xu et al., 2019; Zhou et al., 2020).

Physics-Informed Methods. The physics-informed approach aims to enhance the learning process by incorporating fundamental knowledge and governing physical laws (Karniadakis et al., 2021). In the field of urban computing, this is often achieved either by adding the residuals of physical equations to the loss function as a soft constraint, known as a *learning bias* (Shao et al., 2023; Taghizadeh et al., 2025), or by designing specific network modules or message-passing mechanisms to simulate physical processes, known as an *inductive bias* (Xue; Qu et al., 2023; Gao et al., 2024). However, the former approach often increases training overhead due to the need to compute residuals of partial differential equations (PDEs), while the latter, despite imposing hard constraints, can sacrifice model flexibility. For tasks where the physical processes and rules are relatively well-defined, a more efficient and decoupled pathway is to modify the input data, an approach known as an *observational bias* (Banerjee et al., 2025). For example, Pan et al. (2025) embed principles of traffic flow physics into a step-wise framework for intersection flow prediction by extracting physical performance indicators as input features.

Heterogeneous Graph Methods. In the context of urban physical field prediction, Graph Neural Networks typically simplify the complex urban system into a homogeneous graph (Yu et al., 2024; Zheng & Lu, 2024). However, this simplification limits the model's fidelity and interpretability. Heterogeneous graphs, which consist of multiple types of nodes and edges, can represent the rich semantic relationships in complex systems (Schlichtkrull et al., 2017; Zhang et al., 2019; Zhao et al., 2021). By designing type-aware message-passing mechanisms, Heterogeneous Graph Neural Networks (HGNNs) have achieved success in various tasks, such as quantifying road network homogeneity (Xue et al., 2022), perceiving urban spatial heterogeneity (Xiao et al., 2023), learning urban region representations (Kim & Yoon, 2025), predicting the interactive behaviors of traffic participants (Li et al., 2021), and uncovering the dynamics of building carbon emissions (Yap et al., 2025). Although HGNNs have shown great potential in the field of urban computing, their application in microclimate prediction—specifically, leveraging them to explicitly differentiate between multiple physical processes to enhance model fidelity—remains an unexplored area.

Dynamic Graph Methods. In applications for urban physical field prediction, Graph Neural Networks often rely on a static graph topology to represent the spatial relationships between entities (Mandal & Thakur, 2023; Shao et al., 2024; Xu et al., 2024). However, this assumption conflicts with physical reality, as the scope and intensity of physical processes (e.g., building shading) are determined in real-time by external environmental factors (e.g., solar position). Dynamic Graph Neural Networks (DGNNs) provide a more realistic framework for this problem (Skarding et al., 2021; Zheng et al., 2024). DGNNs have become a mainstream and effective approach for handling other urban tasks with time-varying interactions, particularly in traffic forecasting, demonstrating their potential in the field of urban computing (Zhao et al., 2020; Xie et al., 2020; Bui et al., 2022). In these mainstream applications, the evolution of the graph is typically treated as a data-driven, observational phenomenon (Li et al., 2019; Jin et al., 2020). In physical field prediction tasks, however, the graph topology (e.g., shading relationships) is explicitly reconfigured at each timestep by exogenous physical first principles. This fundamental difference gives rise to the need for a new class of DGNNs capable of learning from graph topologies that are actively reconfigured by physical first principles.

3 PRELIMINARY

Target Variables. The Universal Thermal Climate Index (UTCI) (Jendritzky et al., 2012) and the Physiological Equivalent Temperature (PET) (Matzarakis et al., 1999) represent the isothermal air temperature that would elicit the same physiological stress response. Air Temperature (AT) is the most direct measure of atmospheric heat. Mean Radiant Temperature (MRT) quantifies the radiative heat exchange between the human body and its surrounding surfaces, such as sunlit pavements or shaded building facades. Wind Speed (WS) primarily affects convective heat loss and the efficiency of evaporative cooling from the skin surface. Relative Humidity (RH) determines the efficiency of the body's primary cooling mechanism: sweat evaporation.

ENVI-met model. The data in this paper were generated using the ENVI-met model. ENVI-met is a high-resolution, three-dimensional, non-hydrostatic numerical model widely recognized for simulating surface-plant-air interactions within complex urban structures. The model captures the feedback mechanisms among different urban elements by coupling an atmospheric model with detailed soil and vegetation models. This enables it to accurately simulate how solid boundaries ('hard' boundaries), such as building walls, and porous obstructions ('soft' boundaries), such as vegetation canopies, alter local airflow, temperature, and humidity. The fundamental equations governing these processes are detailed in Appendix A.

Problem Formulation. We model the urban environment by discretizing Geographic Information System (GIS) data into grid cells, where each cell is treated as a node $v \in \mathbb{V}$. The state of the environment is represented by a sequence of dynamic heterogeneous graphs $\{\mathcal{G}_t\}$, where the graph at timestep t is defined as $\mathcal{G}_t = (\mathbb{V}, \mathcal{E}_t, \mathbb{R})$. Here, $\mathbb{V}$ is the static set of nodes, $\mathbb{R}$ is the static set of relation types (e.g., 'covered by shadow from cell'), and $\mathcal{E}_t$ is the set of edges that varies with time. The features of all nodes are collected in a matrix $\boldsymbol{X} \in \mathbb{R}^{|\mathbb{V}|\times 8}$. Additionally, we define $\boldsymbol{u}_t$ as the global context vector at timestep t.

For any one of the six target variables, denoted by k, given a sequence of historical graph observations of length T_{hist}, $\{\mathcal{G}_t\}_{t=t_0-T_{hist}+1}^{t_0}$, and the corresponding sequence of context vectors $\{\boldsymbol{u}_t\}_{t=t_0-T_{hist}+1}^{t_0}$, the model aims to learn a specialized mapping function $\mathcal{F}^{(k)}(\cdot)$ to predict the values of this specific variable for the next T_{pred} timesteps:

$$\left\{\hat{y}_{t_0+1}^{(k)}, \ldots, \hat{y}_{t_0+T_{\text{pred}}}^{(k)}\right\} = \mathcal{F}^{(k)}\left(\{\mathcal{G}_t\}_{t=t_0-T_{\text{hist}}+1}^{t_0}, \{u_t\}_{t=t_0-T_{\text{hist}}+1}^{t_0}, X\right) \tag{1}$$

where $\hat{\boldsymbol{y}}_\tau^{(k)} \in \mathbb{R}^{|\mathbb{V}|}$ is the predicted vector for the target variable k at a future timestep τ.

Relational graph convolutional networks. RGCNs are an extension of GCNs, initially developed for tasks such as link prediction and entity classification. They are specifically designed to handle multi-relational graph data. The core idea is to learn distinct feature transformations for different types of relationships between nodes. The forward-pass update of a single RGCN layer is defined as:

$$\boldsymbol{h}_i^{(l+1)} = \sigma\left(\sum_{r\in\mathbb{R}}\sum_{j\in\mathbb{N}_i^r}\frac{1}{c_{i,r}}\boldsymbol{W}_r^{(l)}\boldsymbol{h}_j^{(l)} + \boldsymbol{W}_0^{(l)}\boldsymbol{h}_i^{(l)}\right) \tag{2}$$

where $\boldsymbol{h}_i^{(l)} \in \mathbb{R}^{d^{(l)}}$ is the hidden state of node v_i in the l-th layer, and $d^{(l)}$ is the dimensionality of the representation at this layer. $\mathbb{N}_i^r$ denotes the set of neighbors of node v_i under relation $r \in \mathbb{R}$. $\boldsymbol{W}_r^{(l)}$ is a learnable, relation-specific weight matrix that allows the model to distinguish between different types of relations, and $\boldsymbol{W}_0^{(l)}$ is the weight matrix for the self-connection. σ represents an element-wise activation function (e.g., PReLU), and $c_{i,r}$ is a problem-specific normalization constant that can either be learned or preset (e.g., $c_{i,r} = \mathbb{N}_i^r$).

4 METHOD

Our proposed UrbanGraph framework consists of two core components: a physics-informed graph representation and a spatio-temporal dynamic relational graph network. To rigorously evaluate the effectiveness of our approach, we first generated a large-scale spatio-temporal dataset through high-fidelity physical simulations, the detailed generation process and parameter configurations of which

are described in Appendix B. Second, to address the challenge that urban systems exhibit high heterogeneity in both spatial and temporal dimensions, we detail our physics-informed graph representation in Section 4.1, which is designed to efficiently capture the underlying physical interactions among different urban elements. Finally, in Section 4.2, we introduce the UrbanGraph architecture, which explicitly leverages the time-varying relationships between urban elements to perform node prediction tasks.

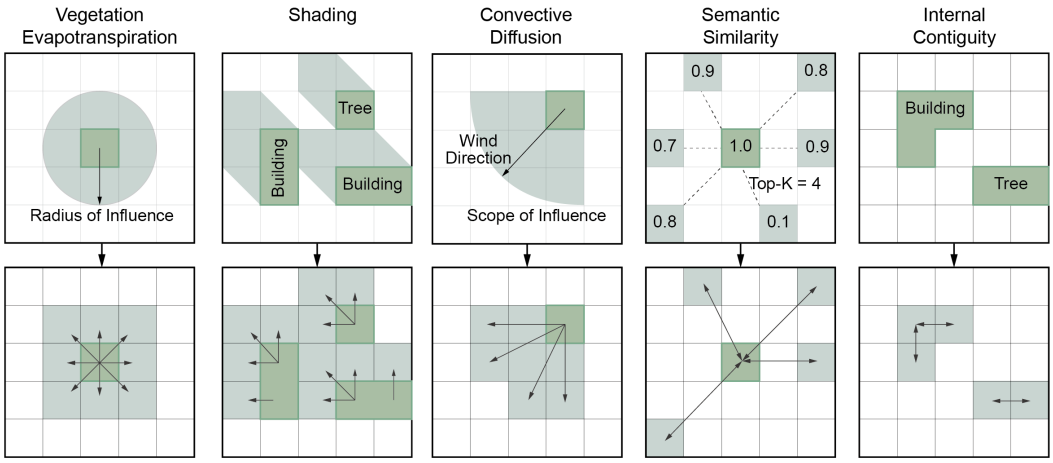

Figure 2: An illustrative overview of the five edge types used in our graph representation. Dynamic edges are derived from physical processes like shadowing and wind, while static edges are based on spatial proximity, feature similarity, and object integrity.

4.1 PHYSICS-INFORMED GRAPH REPRESENTATION

For the graph at any given timestep t, $\mathcal{G}_t = (\mathbb{V}, \mathcal{E}_t, \mathbb{R})$, its edge set $\mathcal{E}_t$ is reconstructed based on the environmental conditions of the current hour. This process is designed to explicitly capture the physical mechanisms that govern the spatial distribution of microclimate factors. The edge set $\mathcal{E}_t$ encodes five distinct types of relationships, which are categorized into two main classes: static and dynamic. Figure 2 provides a visual illustration of the construction mechanisms for these five edge types.

Relational graph convolutional networks. This category of edges represents spatial and semantic relationships that do not change over time. As demonstrated by Yan et al. (2021), non-local spatial functions can significantly influence the assessment of a central region. To capture these non-local interactions between functionally similar nodes, we introduce the following edge type:

SEMANTIC SIMILARITY EDGES. Using the k-Nearest Neighbors (k-NN) algorithm, we construct directed edges from each node to its k nearest neighbors in the normalized static feature space.

INTERNAL CONTIGUITY EDGES. To simulate local energy transfer within large continuous bodies (e.g., building clusters or groups of trees), 'internal nodes' establish connections with their eight immediate neighbors (Moore neighborhood). A node is defined as an 'internal node' if and only if all of its direct neighbors (von Neumann neighborhood) belong to the same object class as itself.

Physics-Informed Dynamic Edges. To explicitly model time-varying physical processes, we introduce three types of dynamic edges whose connections are updated hourly:

SHADING. This edge type is used to model the cooling effect of shadows. A directed edge of type 'shadow' is established from a shading object node v_i (building or tree) with height h_{obj} to a ground node v_j if their Euclidean distance $d(v_i, v_j)$ is less than or equal to the shadow length $L_{shadow,t}$, and the angular deviation falls within a predefined shadow angle width $\Delta\varphi_{shadow}$. The shadow properties are calculated as follows:

$$L_{shadow,t} = h_{obj} / \tan(\theta_{elev,t}) \tag{3}$$

$$\varphi_{shadow,t} = (\varphi_{azimuth,t} + 180^\circ) mod 360^\circ \tag{4}$$

where $\theta_{elev,t}$ is the solar elevation angle at timestep t, $\varphi_{azimuth,t}$ is the solar azimuth angle, and $\varphi_{shadow,t}$ is the principal direction of the shadow projection.

VEGETATION EVAPOTRANSPIRATION. This edge type is designed to represent the local cooling effect of vegetation. A directed edge of type 'Vegetation Evapotranspiration' is established from a tree node v_i to any other node v_j if their Euclidean distance $d(v_i, v_j)$ does not exceed a dynamic radius of influence, $R_{activity,t}$. This radius is calculated based on the global horizontal radiation I_t(Wh/m^2) for the current hour, where R_{base} is a presettable base radius:

$$R_{activity,t} = R_{base} \cdot \text{clip}(I_t/1000, 0.5, 1.2) \tag{5}$$

CONVECTIVE DIFFUSION. To simulate the anisotropic effects of wind-driven convection, an edge of type 'Convective Diffusion' is created from node v_i to v_j. The condition for creating this edge is that their 'effective distance'$d_{eff}(v_i, v_j)$, must be less than or equal to a base local radius, R_{local}. This effective distance is adjusted by a modulation factor, $\alpha_{wind,t}$, which accounts for the wind speed $v_{wind,t}$ and wind direction alignment $\Delta\theta_{wind}$:

$$\alpha_{wind,t} = 1.0 + \lambda_{wind} \cdot \cos(\Delta\theta_{wind}) \cdot (v_{wind,t}/v_{max}) \tag{6}$$

$$d_{eff}(v_i, v_j) = d(v_i, v_j)/\alpha_{wind,t} \leq R_{local} \tag{7}$$

where λ_{wind} is the wind effect intensity coefficient, determining the extent to which wind speed and direction stretch or compress the 'effective connection distance'.v_{max} represents the maximum wind speed observed in the study scenario, ensuring the numerical stability of the model. Detailed parameter configurations for constructing all edge types are provided in Appendix C.

4.2 URBANGRAPH ARCHITECTURE

To effectively process the graph structures from our physics-informed representation, we designed a dynamic and heterogeneous architecture for UrbanGraph. As illustrated in Figure 3, the overall architecture comprises four core components: Feature Encoders, a Spatial Graph Encoder, a Spatio-Temporal Evolution Module, and a Prediction Head.

Feature Encoders and Spatial Graph Encoder. At each timestep t, we employ independent Multi-Layer Perceptrons (MLPs) to encode non-graph dynamic inputs. Graph-level global environmental features $\boldsymbol{u}_t^{env}$ and periodic temporal features $\boldsymbol{u}_t^{time}$ are projected into high-dimensional embedding vectors $\boldsymbol{e}_t^{env}$ and $\boldsymbol{e}_t^{time}$, respectively. For each graph $\mathcal{G}_t$ in the input sequence, we utilize a three-layer RGCN to capture the spatial dependencies defined by the time-varying heterogeneous edges. This module outputs a spatially-informed representation vector $\boldsymbol{h}_{v,t}^{RGCN}$ for each node v.

Spatio-Temporal Evolution Module. This module is responsible for fusing the spatial and global dynamic features and uses a Long Short-Term Memory (LSTM) network to model their temporal evolution. At each prediction timestep t (from t_1 to T_{pred}), we concatenate the spatial representation of a node, $\boldsymbol{h}_{v,t}^{RGCN}$, with the global environmental embedding, $\boldsymbol{e}_t^{env}$, and the temporal embedding, $\boldsymbol{e}_t^{time}$. The resulting concatenated vector is passed through a fusion MLP to generate the input feature for the LSTM layer, $\boldsymbol{x}_{v,t}^{LSTM}$. This is expressed as:

$$\boldsymbol{x}_{v,t}^{LSTM} = \text{MLP}_{fusion}([\boldsymbol{h}_{v,t}^{RGCN} \oplus \boldsymbol{e}_t^{env} \oplus \boldsymbol{e}_t^{time}]) \tag{8}$$

The sequence of fused features is then fed into an LSTM layer to model the temporal dynamics. To provide the model with an effective initial state, an MLP projects the spatial features from the initial graph, $\boldsymbol{h}_{v,t_0}^{RGCN}$, to form the initial hidden state $\boldsymbol{h}_0$. The initial cell state $\boldsymbol{c}_0$ is initialized as a zero vector. This is expressed as:

$$\boldsymbol{h}_0 = \text{MLP}_{h_0}(\boldsymbol{h}_{v,t_0}^{RGCN}) \tag{9}$$

Prediction Head. Finally, a separate MLP decodes the last hidden state of the LSTM, $\boldsymbol{h}_{v,T_{pred}}^{LSTM}$, into a multi-step prediction vector, $\hat{\boldsymbol{y}}_v$. This generates the predictions for all T_{pred} future timesteps at once. This is expressed as:

$$\hat{\boldsymbol{y}}_v = [\hat{y}_{v,1}, \ldots, \hat{y}_{v,Tpred}] = \text{MLP}_{head}(\boldsymbol{h}_{v,T_{pred}}^{LSTM}) \tag{10}$$

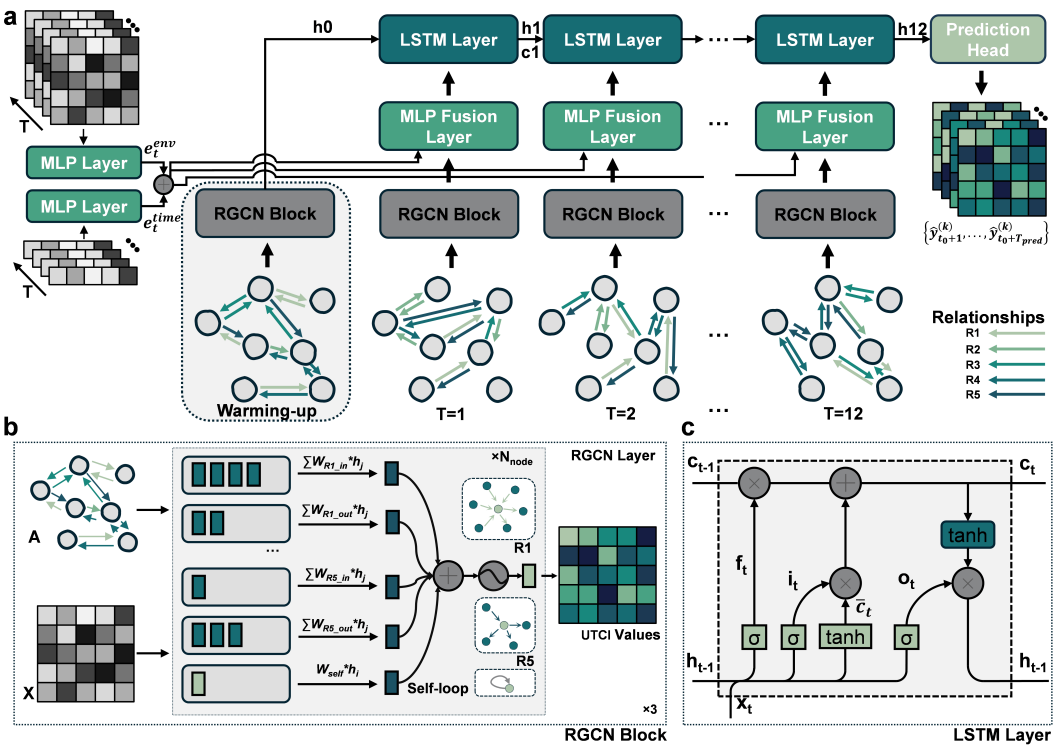

Figure 3: The overall architecture of the UrbanGraph model.(a) The end-to-end framework, which processes historical time-series data, weather context, and a sequence of dynamic heterogeneous graphs. At each timestep, an RGCN Block extracts spatial features from the corresponding graph, which are then fused with temporal features by an MLP Fusion Layer. An LSTM layer propagates the temporal state to the next timestep.(b) The detailed structure of the RGCN Block, which aggregates messages from neighbors across different relation types ($R1 - R5$) and combines them with the node's self-features.(c) The architecture of a standard LSTM layer used for capturing temporal dependencies.

5 Experiments

Dataset. All experiments are conducted on the high-fidelity dataset generated via ENVI-met, as detailed in Appendix B. Our primary task is to evaluate the model's effectiveness on the UTCI prediction task. To further demonstrate the scalability and generalization capability of our physics-informed representation and architecture, we train and evaluate models on all six target variables. The temporal graph structures are constructed following the physical principles outlined in Section 4.1. The resulting spatio-temporal graph sequence data is split into training (70%), validation (20%), and testing (10%) sets.

Baseline Models. To comprehensively evaluate our model's performance, we construct a series of strong baseline models by replacing key components of our proposed architecture. While ensuring a relatively fair comparison, we independently optimize the hyperparameters for each architecture. (i)Non-Graph Model.We replace the RGCN module with a Conditional GAN (CGAN) (Isola et al., 2018) to evaluate the utility of an explicit graph structure. (ii)Homogeneous Graph Models. We replace the RGCN module with GCN and GINE to assess the importance of modeling heterogeneous relationships. (iii)Generative Graph Models. We replace the RGCN with generative methods, including a Graph Autoencoder (GAE) and a Graph GAN (GGAN), where the CNN layers in a standard GAN are replaced with RGCN layers. (iv)Temporal Variants.We replace the LSTM module with a GRU and a Transformer encoder to analyze the impact of different temporal modeling components.

Evaluation Metrics. The predictive performance of the models is evaluated using three standard regression metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). To assess computational cost, we record the number of floating-point operations (FLOPs), training time (seconds per epoch), and inference speed (inferences per second). All models are trained using the Adam optimizer with a Mean Squared Error (MSE) loss function. For the GAE, the loss function is augmented with a KL divergence term; for the CGAN, a binary cross-entropy loss is added. To ensure robust training, we employ the ReduceLROnPlateau learning rate scheduling strategy and an early stopping mechanism.

Model Settings. In the main comparative analysis, our proposed spatio-temporal heterogeneous model is configured with a learning rate of 0.001, a batch size of 8, a hidden dimension of 128 for all layers, a 3-layer RGCN encoder, and a 1-layer LSTM. It uses a multi-head prediction architecture, and all models are run for 3 independent trials. For the subsequent ablation studies and sensitivity analyses, we use a model with hyperparameters optimized by Optuna (Akiba et al., 2019), featuring a hidden dimension of 384 and a single prediction head. More detailed hyperparameter settings are available in Appendix D. All experiments were conducted on a single NVIDIA L4 GPU.

6 RESULT

This section presents a comprehensive experimental evaluation of our proposed UrbanGraph framework. We begin in Section 6.1 by comparing its performance against various baseline models to validate its effectiveness. Subsequently, in Section 6.2, we conduct key ablation studies to quantify the contributions of our model's core components. Further detailed analyses—including additional ablation studies, hyperparameter sensitivity, and a computational performance evaluation—are provided in Appendix E.

6.1 MODEL PERFORMANCE

As shown in Table 1, our proposed UrbanGraph achieves the best performance across all evaluation metrics. On the test set, the model achieves the highest average R^2 of 0.8542 and the lowest RMSE of 1.0535, outperforming all baseline models. Compared to all baselines, it improves prediction accuracy by up to 10.8% (in R^2) and enhances computational efficiency by 17.0% (in FLOPs). This highlights the effectiveness of the proposed method in efficiently capturing complex spatio-temporal dependencies.

Table 1: Performance and efficiency comparison of different model architectures on the test set.

Model	Flops	Test		Time Cost	
		Avg R^2	Avg RMSE	Training (epoch/s)	Inference/s
CGAN-LSTM	1.10×10^{10}	$0.7712 \pm .0369$	$1.3450 \pm .1175$	15.3252 ± 1.0999	$1.5558 \pm .1951$
GCN-LSTM	8.28×10^{9}	$0.8347 \pm .0039$	$1.1327 \pm .0433$	28.5321 ± 2.8358	$2.8619 \pm .4516$
GINE-LSTM	8.80×10^{9}	$0.8087 \pm .0226$	$1.2045 \pm .0294$	32.3169 ± 1.4643	$3.1731 \pm .2325$
GAE-LSTM	1.05×10^{10}	$0.8494 \pm .0036$	$1.0687 \pm .0269$	36.7376 ± 3.2079	$3.6022 \pm .4504$
GGAN-LSTM	9.44×10^{9}	$0.8415 \pm .0034$	$1.0981 \pm .0406$	42.4678 ± 3.1537	$2.6488 \pm .4073$
RGCN-GRU	7.12×10^{9}	$0.8483 \pm .0035$	$1.0682 \pm .0380$	20.8096 ± 1.3612	$2.1640 \pm .2133$
RGCN-Transformer	5.09×10^{10}	$0.8465 \pm .0065$	$1.0791 \pm .0253$	$37.6463 \pm .8325$	$3.3345 \pm .1482$
URBANGRAPH	9.13×10^{9}	$\mathbf{0.8542 \pm .0044}$	$\mathbf{1.0535 \pm .0338}$	$\mathbf{24.4823 \pm 0.9323}$	$\mathbf{2.6914 \pm .1404}$

The convergence curve (Figure 4a) confirms the stability of the model's training process. Moreover, the hour-by-hour error analysis (Figure 4b) shows that our method consistently maintains the lowest RMSE throughout the entire 12-hour prediction horizon. It demonstrates strong robustness against error accumulation, particularly during afternoon hours (e.g., 14:00 and 17:00) when climate fluctuations are more pronounced.

In addition to these quantitative metrics, we provide qualitative visualizations of the predicted heat maps in Appendix F, which intuitively demonstrate the model's ability to capture fine-grained spatial distributions.

6.2 ABLATION ANALYSIS

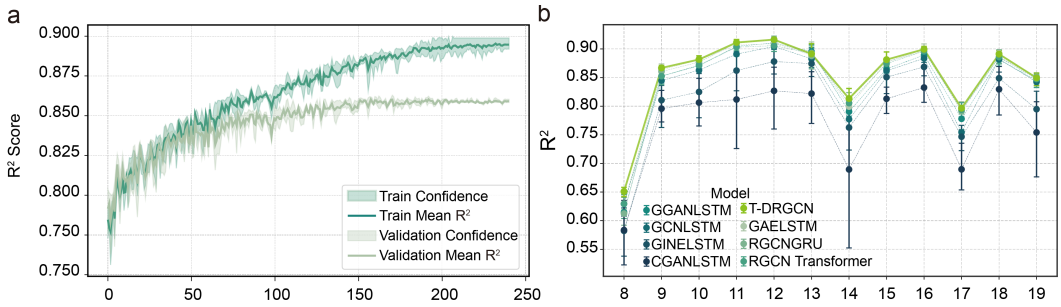

Figure 4: Model performance analysis. (a) Training and validation R^2 convergence curves for the UrbanGraph model. Shaded areas represent the confidence interval. (b) Hour-by-hour R^2 comparison between UrbanGraph and baselines on the test set, with error bars indicating standard deviation.

Heterogeneous Graph Mechanism. To validate the importance of modeling diverse physical interactions with distinct relation types, we compare our full model (Base), which uses a heterogeneous graph (RGCN), against a variant that simplifies the graph to be homogeneous (GCN). The results, shown in Table 2a, reveal a significant performance degradation when heterogeneity is removed, with the R^2 score dropping from 0.8629 to 0.8347. This underscores the necessity of using a heterogeneous framework to differentiate between various physical processes.

Table 2: Ablation studies for key mechanisms.

(a) Heterogeneous.

Model	R^2	MSE
Base	**0.8629**	**1.0976**
Homo	0.8336	1.4275

(b) Dynamic.

Model	R^2	MSE
Base	**0.8629**	**1.0976**
Static	0.8057	1.6678

Dynamic Graph Mechanism. To validate the effectiveness of the dynamic graph mechanism, we compare our model with a variant that uses a static graph (i.e., the same graph structure is shared across all timesteps). As shown in Table 2b, disabling the dynamic mechanism leads to a significant performance drop in the model (Static), with the R^2 score decreasing from 0.8629 to 0.8057. This highlights that explicitly modeling the temporal evolution of spatial interactions is crucial for this prediction task.

Further ablation studies analyzing other key components—such as the contribution of individual edge types, various prediction head architectures, feature fusion strategies, and the effects of explicit edge features—are detailed in Appendix E.

7 Conclusion

In this paper, we proposed UrbanGraph, a physics-informed dynamic heterogeneous graph framework for solving urban dynamic heterogeneous graph computing tasks. We tested it on the problem of microclimate prediction, where UrbanGraph achieved the best performance compared to all baselines. It improves prediction accuracy by up to 10.8% (in R^2) and computational efficiency by 17.0% (in FLOPs), with the heterogeneous and dynamic graph mechanisms contributing gains of 3.5% and 7.1%, respectively. Furthermore, the UMC4/12 dataset, which we constructed and released, serves as the first high-resolution benchmark in this field and will help accelerate the development and fair comparison of new algorithms in the future. In summary, our work advances the application of data-driven methods in the field of urban physical field prediction and points to a promising direction for future research.

Limitation and Future Work. Our work explicitly encodes predefined physical processes (i.e., prior knowledge) into the graph topology. While this has shown performance advantages, it may oversimplify the real physical processes, as it might overlook latent relationships present in the data that we have not yet modeled or are unknown. Therefore, a key direction for future research is to explore adaptive graph learning methods. The core objective is to design a framework that can automatically learn and optimize the graph structure from data in an end-to-end manner. Such a framework could not only surpass existing frameworks on prediction tasks but also uncover unknown interaction patterns among urban entities, providing new insights for urban science.

Reproducibility Statement

To ensure reproducibility, our code and the UMC4/12 dataset are publicly available at https://anonymous.4open.science/r/UrbanGraph/.

Acknowledgments

The authors acknowledge the use of a large language model for assistance with language editing and improving the clarity of this manuscript.

References

Reem Abd Elraouf, Ashraf Elmokadem, Naglaa Megahed, Osama Abo Eleinen, and Sara Eltarabily. The impact of urban geometry on outdoor thermal comfort in a hot-humid climate. *Building and Environment*, 225:109632, November 2022. ISSN 0360-1323. doi: 10.1016/j.buildenv.2022.109632. URL https://www.sciencedirect.com/science/article/pii/S0360132322008629.

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework, July 2019. URL http://arxiv.org/abs/1907.10902. arXiv:1907.10902 [cs].

Meryem El Alaoui, Laila Ouazzani Chahidi, Mohamed Rougui, Abdellah Mechaqrane, and Senhaji Allal. Evaluation of CFD and machine learning methods on predicting greenhouse microclimate parameters with the assessment of seasonality impact on machine learning performance. *Scientific African*, 19:e01578, March 2023. ISSN 2468-2276. doi: 10.1016/j.sciaf.2023.e01578. URL https://www.sciencedirect.com/science/article/pii/S2468227623000376.

Elanchezhian Arulmozhi, Jayanta Kumar Basak, Thavisack Sihalath, Jaesung Park, Hyeon Tae Kim, and Byeong Eun Moon. Machine Learning-Based Microclimate Model for Indoor Air Temperature and Relative Humidity Prediction in a Swine Building. *Animals*, 11(1):222, January 2021. ISSN 2076-2615. doi: 10.3390/ani11010222. URL https://www.mdpi.com/2076-2615/11/1/222. Publisher: Multidisciplinary Digital Publishing Institute.

Chayan Banerjee, Kien Nguyen, Clinton Fookes, and Karniadakis George. Physics-Informed Computer Vision: A Review and Perspectives. *ACM Computing Surveys*, 57(1):1–38, January 2025. ISSN 0360-0300, 1557-7341. doi: 10.1145/3689037. URL https://dl.acm.org/doi/10.1145/3689037.

Guilhardo Barros Moreira de Carvalho and Luiz Bueno da Silva. The microclimate implications of urban form applying computer simulation: systematic literature review. *Environment, Development and Sustainability*, 26(10):24687–24726, October 2024. ISSN 1573-2975. doi: 10.1007/s10668-023-03737-5. URL https://doi.org/10.1007/s10668-023-03737-5.

Michael Bruse and Heribert Fleer. Simulating surface–plant–air interactions inside urban environments with a three dimensional numerical model. *Environmental Modelling & Software*, 13(3):373–384, October 1998. ISSN 1364-8152. doi: 10.1016/S1364-8152(98)00042-5. URL https://www.sciencedirect.com/science/article/pii/S1364815298000425.

Khac-Hoai Nam Bui, Jiho Cho, and Hongsuk Yi. Spatial-temporal graph neural network for traffic forecasting: An overview and open research issues. *Applied Intelligence*, 52(3):2763–2774, February 2022. ISSN 1573-7497. doi: 10.1007/s10489-021-02587-w. URL https://doi.org/10.1007/s10489-021-02587-w.

Anna Carter, Michael Kearney, Nicola Mitchell, Stephen Hartley, Warren Porter, and Nicola Nelson. Modelling the soil microclimate: does the spatial or temporal resolution of input parameters matter? *Frontiers of Biogeography*, 7(4), January 2016. ISSN 1948-6596. doi: 10.21425/F5FBG27849. URL https://escholarship.org/uc/item/9sq12044.

Andrew M. Coutts, Nigel J. Tapper, Jason Beringer, Margaret Loughnan, and Matthias Demuzere. Watering our cities: The capacity for Water Sensitive Urban Design to support urban cooling and improve human thermal comfort in the Australian context. *Progress in Physical Geography: Earth and Environment*, 37(1):2–28, February 2013. ISSN 0309-1333. doi: 10.1177/0309133312461032. URL https://doi.org/10.1177/0309133312461032. Publisher: SAGE Publications Ltd.

Loyde Vieira de Abreu-Harbich, Lucila Chebel Labaki, and Andreas Matzarakis. Effect of tree planting design and tree species on human thermal comfort in the tropics. *Landscape and Urban Planning*, 138:99–109, June 2015. ISSN 0169-2046. doi: 10.1016/j.landurbplan.2015.02.008. URL https://www.sciencedirect.com/science/article/pii/S0169204615000390.

Kunihiko Fujiwara, Maxim Khomiakov, Winston Yap, Marcel Ignatius, and Filip Biljecki. Microclimate Vision: Multimodal prediction of climatic parameters using street-level and satellite imagery. *Sustainable Cities and Society*, 114:105733, November 2024. ISSN 22106707. doi: 10.1016/j.scs.2024.105733. URL https://linkinghub.elsevier.com/retrieve/pii/S2210670724005584.

Huanxiang Gao, Gang Hu, Dongqin Zhang, Wenjun Jiang, K. T. Tse, K. C. S. Kwok, and Ahsan Kareem. Urban wind field prediction based on sparse sensors and physics-informed graph-assisted auto-encoder. *Computer-Aided Civil and Infrastructure Engineering*, 39(10):1409–1430, 2024. ISSN 1467-8667. doi: 10.1111/mice.13147. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.13147. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mice.13147.

Leon Yan-Feng Gaw, Alex Thiam Koon Yee, and Daniel Rex Richards. A High-Resolution Map of Singapore's Terrestrial Ecosystems. *Data*, 4(3):116, August 2019. ISSN 2306-5729. doi: 10.3390/data4030116. URL https://www.mdpi.com/2306-5729/4/3/116.

Luke Grant, Inne Vanderkelen, Lukas Gudmundsson, Erich Fischer, Sonia I. Seneviratne, and Wim Thiery. Global emergence of unprecedented lifetime exposure to climate extremes. *Nature*, 641 (8062):374–379, May 2025. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-025-08907-1. URL https://www.nature.com/articles/s41586-025-08907-1. Publisher: Springer Science and Business Media LLC.

M. Haklay and P. Weber. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*, 7(4):12–18, October 2008. ISSN 1536-1268. doi: 10.1109/MPRV.2008.80. URL http://ieeexplore.ieee.org/document/4653466/.

Mengyuan He, Hong Liu, Zhaosong Fang, Bo He, and Baizhan Li. High-temperature and thermal radiation affecting human thermal comfort and physiological responses: An experimental study. *Journal of Building Engineering*, 86:108815, June 2024. ISSN 2352-7102. doi: 10.1016/j.jobe.2024.108815. URL https://www.sciencedirect.com/science/article/pii/S2352710224003838.

Yeonsook Heo, Gigih R. Setyantho, and Tageui Hong. Toward a new paradigm for urban climate modelling: challenges and opportunities. *Journal of Building Performance Simulation*, 0(0):1–8. ISSN 1940-1493. doi: 10.1080/19401493.2025.2540925. URL https://doi.org/10.1080/19401493.2025.2540925. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/19401493.2025.2540925.

M. Akif Irmak, Sevgi Yilmaz, Doğan Dursun, M. Akif Irmak, Sevgi Yilmaz, and Doğan Dursun. Effect of different pavements on human thermal comfort conditions. *Atmósfera*, 30(4):355–366, 2017. ISSN 0187-6236. doi: 10.20937/atm.2017.30.04.06. URL http://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S0187-62362017000400355&lng=es&nrm=iso&tlng=en. Publisher: Instituto de Ciencias de la Atmósfera y Cambio Climático, UNAM.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks, November 2018. URL http://arxiv.org/abs/1611.07004. arXiv:1611.07004 [cs].

Gerd Jendritzky, Richard de Dear, and George Havenith. UTCI—Why another thermal index? *International Journal of Biometeorology*, 56(3):421–428, May 2012. ISSN 1432-1254. doi: 10.1007/s00484-011-0513-7. URL https://doi.org/10.1007/s00484-011-0513-7.

Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs, October 2020. URL http://arxiv.org/abs/1904.05530. arXiv:1904.05530 [cs].

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, June 2021. ISSN 2522-5820. doi: 10.1038/s42254-021-00314-5. URL https://www.nature.com/articles/s42254-021-00314-5. Publisher: Nature Publishing Group.

Patrick Kastner and Timur Dogan. A GAN-Based Surrogate Model for Instantaneous Urban Wind Flow Prediction. *Building and Environment*, 242:110384, August 2023. ISSN 03601323. doi: 10.1016/j.buildenv.2023.110384. URL https://linkinghub.elsevier.com/retrieve/pii/S0360132323004110.

Julia Kemppinen, Jonas J. Lembrechts, Koenraad Van Meerbeek, et al. Microclimate, an important part of ecology and biogeography. *Global Ecology and Biogeography*, 33(6):e13834, 2024. ISSN 1466-8238. doi: 10.1111/geb.13834. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13834. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/geb.13834.

Ji Yeon Kim, Chae Yeon Park, Dong Kun Lee, Seok Hwan Yun, Jung Hee Hyun, and Eun Sub Kim. The cooling effect of trees in high-rise building complexes in relation to spatial distance from buildings. *Sustainable Cities and Society*, 114:105737, November 2024. ISSN 2210-6707. doi: 10.1016/j.scs.2024.105737. URL https://www.sciencedirect.com/science/article/pii/S2210670724005626.

Namwoo Kim and Yoonjin Yoon. Effective Urban Region Representation Learning Using Heterogeneous Urban Graph Attention Network (HUGAT). *IEEE Access*, 13:102602–102612, 2025. ISSN 2169-3536. doi: 10.1109/ACCESS.2025.3577202. URL https://ieeexplore.ieee.org/abstract/document/11027114.

Thomas N. Kipf and Max Welling. Variational Graph Auto-Encoders, November 2016. URL http://arxiv.org/abs/1611.07308. arXiv:1611.07308 [stat].

Peeyush Kumar, Ranveer Chandra, Chetan Bansal, Shivkumar Kalyanaraman, Tanuja Ganu, and Michael Grant. Micro-climate Prediction - Multi Scale Encoder-decoder based Deep Learning Framework. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3128–3138, Virtual Event Singapore, August 2021. ACM. ISBN 978-1-4503-8332-5. doi: 10.1145/3447548.3467173. URL https://dl.acm.org/doi/10.1145/3447548.3467173.

Jia Li, Zhichao Han, Hong Cheng, Jiao Su, Pengyun Wang, Jianfeng Zhang, and Lujia Pan. Predicting Path Failure In Time-Evolving Graphs, May 2019. URL http://arxiv.org/abs/1905.03994. arXiv:1905.03994 [cs].

Zirui Li, Jianwei Gong, Chao Lu, and Yangtian Yi. Interactive Behavior Prediction for Heterogeneous Traffic Participants in the Urban Road: A Graph-Neural-Network-Based Multitask Learning Framework. *IEEE/ASME Transactions on Mechatronics*, 26(3):1339–1349, June 2021. ISSN 1941-014X. doi: 10.1109/TMECH.2021.3073736. URL https://ieeexplore.ieee.org/document/9406384.

Zhixin Liu, Wenwen Cheng, C.Y. Jim, Tobi Eniolu Morakinyo, Yuan Shi, and Edward Ng. Heat mitigation benefits of urban green and blue infrastructures: A systematic review of modeling techniques, validation and scenario simulation in ENVI-met V4. *Building and Environment*, 200:107939, August 2021. ISSN 03601323. doi: 10.1016/j.buildenv.2021.107939. URL https://linkinghub.elsevier.com/retrieve/pii/S0360132321003437.

Subhojit Mandal and Mainak Thakur. A city-based PM2.5 forecasting framework using Spatially Attentive Cluster-based Graph Neural Network model. *Journal of Cleaner Production*, 405:137036, June 2023. ISSN 09596526. doi: 10.1016/j.jclepro.2023.137036. URL https://linkinghub.elsevier.com/retrieve/pii/S0959652623011940.

Andreas Matzarakis, H. Mayer, and M.G. Iziomon. Applications of a universal thermal index: physiological equivalent temperature. *International Journal of Biometeorology*, 43(2):76–84, October 1999. ISSN 1432-1254. doi: 10.1007/s004840050119. URL https://doi.org/10.1007/s004840050119.

Yuyan Annie Pan, Fuliang Li, Anran Li, Zhiqiang Niu, and Zhen Liu. Urban intersection traffic flow prediction: A physics-guided stepwise framework utilizing spatio-temporal graph neural network algorithms. *Multimodal Transportation*, 4(2):100207, June 2025. ISSN 2772-5863. doi: 10.1016/j.multra.2025.100207. URL https://www.sciencedirect.com/science/article/pii/S2772586325000218.

Haohao Qu, Haoxuan Kuang, Jun Li, and Linlin You. A physics-informed and attention-based graph learning approach for regional electric vehicle charging demand prediction, November 2023. URL http://arxiv.org/abs/2309.05259. arXiv:2309.05259 [cs].

Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling Relational Data with Graph Convolutional Networks, October 2017. URL http://arxiv.org/abs/1703.06103. arXiv:1703.06103 [stat].

Xuqiang Shao, Zhijian Liu, Siqi Zhang, Zijia Zhao, and Chenxing Hu. PIGNN-CFD: A physics-informed graph neural network for rapid predicting urban wind field defined on unstructured mesh. *Building and Environment*, 232:110056, March 2023. ISSN 03601323. doi: 10.1016/j.buildenv.2023.110056. URL https://linkinghub.elsevier.com/retrieve/pii/S0360132323000835.

Xuqiang Shao, Siqi Zhang, Xiaofan Liu, Zhijian Liu, and Jiancai Huang. Rapid prediction for the transient dispersion of leaked airborne pollutant in urban environment based on graph neural networks. *Journal of Hazardous Materials*, 478:135517, October 2024. ISSN 03043894. doi: 10.1016/j.jhazmat.2024.135517. URL https://linkinghub.elsevier.com/retrieve/pii/S030438942402096X.

Joakim Skarding, Bogdan Gabrys, and Katarzyna Musial. Foundations and modelling of dynamic networks using Dynamic Graph Neural Networks: A survey. *IEEE Access*, 9:79143–79168, 2021. ISSN 2169-3536. doi: 10.1109/ACCESS.2021.3082932. URL http://arxiv.org/abs/2005.07496. arXiv:2005.07496 [cs].

Mehdi Taghizadeh, Zanko Zandsalimi, Mohammad Amin Nabian, Majid Shafiee-Jood, and Negin Alemazkoor. Interpretable physics-informed graph neural networks for flood forecasting. *Computer-Aided Civil and Infrastructure Engineering*, 40(18):2629–2649, 2025. ISSN 1467-8667. doi: 10.1111/mice.13484. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.13484. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mice.13484.

Jamie Tolan, Hung-I Yang, Benjamin Nosarzewski, Guillaume Couairon, Huy V. Vo, John Brandt, Justine Spore, Sayantan Majumdar, Daniel Haziza, Janaki Vamaraju, Theo Moutakanni, Piotr Bojanowski, Tracy Johns, Brian White, Tobias Tiecke, and Camille Couprie. Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment*, 300:113888, January 2024. ISSN 00344257. doi: 10.1016/j.rse.2023.113888. URL https://linkinghub.elsevier.com/retrieve/pii/S003442572300439X.

Y. Toparlar, B. Blocken, B. Maiheu, and G.J.F. Van Heijst. A review on the CFD analysis of urban microclimate. *Renewable and Sustainable Energy Reviews*, 80:1613–1640, December 2017. ISSN 13640321. doi: 10.1016/j.rser.2017.05.248. URL https://linkinghub.elsevier.com/retrieve/pii/S1364032117308924.

S. Tsoka, A. Tsikaloudaki, and T. Theodosiou. Analyzing the ENVI-met microclimate model's performance and assessing cool materials and urban vegetation applications–A review. *Sustainable Cities and Society*, 43:55–76, November 2018. ISSN 22106707. doi: 10.1016/j.scs.2018.08.009. URL https://linkinghub.elsevier.com/retrieve/pii/S2210670718307649.

Congxi Xiao, Jingbo Zhou, Jizhou Huang, Tong Xu, and Hui Xiong. Spatial Heterophily Aware Graph Neural Networks. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2752–2763, August 2023. doi: 10.1145/3580305.3599510. URL http://arxiv.org/abs/2306.12139. arXiv:2306.12139 [cs].

Yi Xie, Yun Xiong, and Yangyong Zhu. SAST-GNN: A Self-Attention Based Spatio-Temporal Graph Neural Network for Traffic Prediction. In Yunmook Nah, Bin Cui, Sang-Won Lee, Jeffrey Xu Yu, Yang-Sae Moon, and Steven Euijong Whang (eds.), *Database Systems for Advanced Applications*, pp. 707–714, Cham, 2020. Springer International Publishing. ISBN 978-3-030-59410-7. doi: 10.1007/978-3-030-59410-7_49.

Hongbin Xu, Siyi Zhang, and Chong Wu. Revealing the Impact of Urban Land Use Patterns on Land Surface Temperature Through Graph Attention Networks, 2024. URL https://www.ssrn.com/abstract=5074905.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How Powerful are Graph Neural Networks?, February 2019. URL http://arxiv.org/abs/1810.00826. arXiv:1810.00826 [cs].

Jiawei Xue. PHYSICS-INFORMED GRAPH LEARNING IN URBAN TRAFFIC NETWORKS.

Jiawei Xue, Nan Jiang, Senwei Liang, Qiyuan Pang, Takahiro Yabe, Satish V. Ukkusuri, and Jianzhu Ma. Quantifying the spatial homogeneity of urban road networks via graph neural networks. *Nature Machine Intelligence*, 4(3):246–257, March 2022. ISSN 2522-5839. doi: 10.1038/s42256-022-00462-y. URL https://www.nature.com/articles/s42256-022-00462-y. Publisher: Nature Publishing Group.

Longxu Yan, De Wang, Shangwu Zhang, and Carlo Ratti. Understanding urban centers in Shanghai with big data: Local and non-local function perspectives. *Cities*, 113:103156, June 2021. ISSN 0264-2751. doi: 10.1016/j.cities.2021.103156. URL https://www.sciencedirect.com/science/article/pii/S0264275121000548.

Winston Yap, Abraham Noah Wu, Clayton Miller, and Filip Biljecki. Revealing building operating carbon dynamics for multiple cities. *Nature Sustainability*, pp. 1–12, August 2025. ISSN 2398-9629. doi: 10.1038/s41893-025-01615-8. URL https://www.nature.com/articles/s41893-025-01615-8. Publisher: Nature Publishing Group.

Yin Yu, Peiyuan Li, Daning Huang, and Ashish Sharma. Street-level temperature estimation using graph neural networks: Performance, feature embedding and interpretability. *Urban Climate*, 56:102003, July 2024. ISSN 22120955. doi: 10.1016/j.uclim.2024.102003. URL https://linkinghub.elsevier.com/retrieve/pii/S2212095524001998.

Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V. Chawla. Heterogeneous Graph Neural Network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pp. 793–803, New York, NY, USA, July 2019. Association for Computing Machinery. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330961. URL https://dl.acm.org/doi/10.1145/3292500.3330961.

Jianan Zhao, Xiao Wang, Chuan Shi, Binbin Hu, Guojie Song, and Yanfang Ye. Heterogeneous Graph Structure Learning for Graph Neural Networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5):4697–4705, May 2021. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v35i5.16600. URL https://ojs.aaai.org/index.php/AAAI/article/view/16600.

Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Transactions*

on Intelligent Transportation Systems, 21(9):3848–3858, September 2020. ISSN 1524-9050, 1558-0016. doi: 10.1109/TITS.2019.2935152. URL https://ieeexplore.ieee.org/document/8809901/.

Lang Zheng and Weisheng Lu. Urban micro-scale street thermal comfort prediction using a 'graph attention network' model. *Building and Environment*, 262:111780, August 2024. ISSN 03601323. doi: 10.1016/j.buildenv.2024.111780. URL https://linkinghub.elsevier.com/retrieve/pii/S036013232400622X.

Yanping Zheng, Lu Yi, and Zhewei Wei. A survey of dynamic graph neural networks, April 2024. URL http://arxiv.org/abs/2404.18211. arXiv:2404.18211 [cs].

Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81, January 2020. ISSN 2666-6510. doi: 10.1016/j.aiopen.2021.01.001. URL https://www.sciencedirect.com/science/article/pii/S2666651021000012.

A Key Physical Equations in ENVI-met

This appendix outlines the key physical equations within the ENVI-met model (Bruse & Fleer, 1998) used to generate the dataset for this study.

A.1 Mean Air Flow

The model describes three-dimensional turbulence by solving the non-hydrostatic, incompressible Navier-Stokes equations. The fundamental equations for the mean wind velocity components u,v,w are as follows:

$$\frac{\partial u}{\partial t} + u_i \frac{\partial u}{\partial x_i} = -\frac{\partial p'}{\partial x} + K_m \left(\frac{\partial^2 u}{\partial x_i^2} \right) + f(v - v_g) - S_u \tag{11}$$

$$\frac{\partial v}{\partial t} + u_i \frac{\partial v}{\partial x_i} = -\frac{\partial p'}{\partial y} + K_m \left(\frac{\partial^2 v}{\partial x_i^2} \right) - f(u - u_g) - S_v \tag{12}$$

$$\frac{\partial w}{\partial t} + u_i \frac{\partial w}{\partial x_i} = -\frac{\partial p'}{\partial z} + K_m \left(\frac{\partial^2 w}{\partial x_i^2} \right) + g \frac{\theta(z)}{\theta_{ref}(z)} - S_w \tag{13}$$

where $p'$is the local pressure perturbation,θ is the potential temperature, K_m is the turbulent diffusivity for momentum, f is the Coriolis parameter, and $S_{u(i)}$ are the momentum source/sink terms induced by elements such as vegetation.

A.2 Temperature and Humidity

The distribution of potential temperature θ and specific humidity q in the atmosphere is described by the advection-diffusion equations, which include internal source/sink terms:

$$\frac{\partial \theta}{\partial t} + u_i \frac{\partial \theta}{\partial x_i} = K_h \left(\frac{\partial^2 \theta}{\partial x_i^2} \right) + Q_h \tag{14}$$

$$\frac{\partial q}{\partial t} + u_i \frac{\partial q}{\partial x_i} = K_q \left(\frac{\partial^2 q}{\partial x_i^2} \right) + Q_q \tag{15}$$

where $K_h$and K_q are the turbulent exchange coefficients for heat and moisture, respectively. Q_h and Q_q are the source/sink terms that couple the heat and moisture exchange processes at the surface and with vegetation.

A.3 RADIATIVE FLUXES

The model solves the energy balance for surfaces and walls by calculating the net shortwave radiation, $R_{sw,net}$, and the net longwave radiation,$R_{lw,net}$. The shortwave radiation flux at any point, $R_{sw}(z)$, consists of direct and diffuse radiation, and accounts for the shading effects of buildings and vegetation:

$$R_{sw}(z) = \sigma_{sw,dir}(z)R^0_{sw,dir} + \sigma_{sw,dif}(z)\sigma_{svf}(z)R^0_{sw,dif} + (1 - \sigma_{svf}(z))R^0_{sw,dif}\bar{\alpha} \quad (16)$$

where the R^0 terms represent the incoming radiation at the top of the model, and the σ coefficients are reduction factors ranging from 0 to 1 that quantify the effects of direct radiation $\sigma_{sw,dir}$, diffuse radiation $\sigma_{sw,dif}$, and the sky view factor σ_{svf}.

For the complete set of model equations, parameterization schemes, and numerical solution methods, please refer to the original publication.

B HIGH-RESOLUTION SPATIO-TEMPORAL DATASET FOR MICROCLIMATE AND THERMAL COMFORT

We constructed the UMC4/12 dataset based on public geospatial data and the ENVI-met model. We selected a typical extreme heat day as the basis for our simulations, using the standard meteorological year data (EPW) from Singapore Changi Airport. To ensure morphological diversity in the dataset, we employed a stratified sampling strategy to select 11 representative 1 km² sites across Singapore. The stratification was based on key urban morphology metrics, and the sample pool covers a wide range of urban typologies, from ultra-high-density commercial districts to mature residential areas with large parks (see Appendix Table A1). The metrics include Average Building Height (Avg.BH), Green Space Ratio (GSR), and Building Coverage Ratio (BCR).

Table A1: Distribution of morphological and material properties for the 11 selected 1km² sites in Singapore.

Data Index	Avg.BH(m)	GSR	BCR	Pavement%	Smashed Brick%	Loamy Soil%	Deep Water%
1	13.36	0.021	0.078	0.520	0.095	0.367	0.019
2	19.76	0.055	0.155	0.734	0.062	0.181	0.023
3	12.86	0.255	0.219	0.487	0.000	0.471	0.043
4	23.97	0.184	0.217	0.630	0.043	0.316	0.010
5	10.11	0.116	0.235	0.643	0.026	0.302	0.029
6	12.01	0.429	0.126	0.260	0.020	0.718	0.002
7	28.14	0.165	0.242	0.671	0.023	0.286	0.020
8	13.86	0.209	0.338	0.733	0.036	0.220	0.011
9	33.73	0.198	0.108	0.297	0.050	0.554	0.098
10	19.81	0.105	0.109	0.291	0.023	0.222	0.464
11	16.06	0.444	0.128	0.211	0.062	0.654	0.072

We built the 3D model input files for the ENVI-met simulations by integrating multiple public geospatial data sources. Specifically, we resampled and performed 3D voxelization on building footprints and heights from OpenStreetMap (Haklay & Weber, 2008), a high-resolution land cover classification map (Gaw et al., 2019), and an ultra-high-resolution canopy height map (Tolan et al., 2024). This process generated ENVI-met input files (.INX) with a horizontal resolution of 4 meters and a vertical resolution of 3 meters. To ensure high fidelity, we assigned realistic material properties to different surfaces and building boundaries, and specified corresponding tree species for vegetation of varying heights. The detailed material assignments and parameters are provided in Appendix Table A2 and A3. The simulation period covered the hours from 08:00 to 19:00, when urban heat effects are most significant.

Figure A1 provides a visualization of the primary input data layers—tree height, land cover type, and building height—for two representative sites, illustrating the morphological diversity within the UMC4/12 dataset.

Following the ENVI-met simulation, we generated high-resolution spatio-temporal data for the six target variables. Figure A2 illustrates the simulation output for one of the urban blocks, displaying the evolution of all six variables over the course of the day.

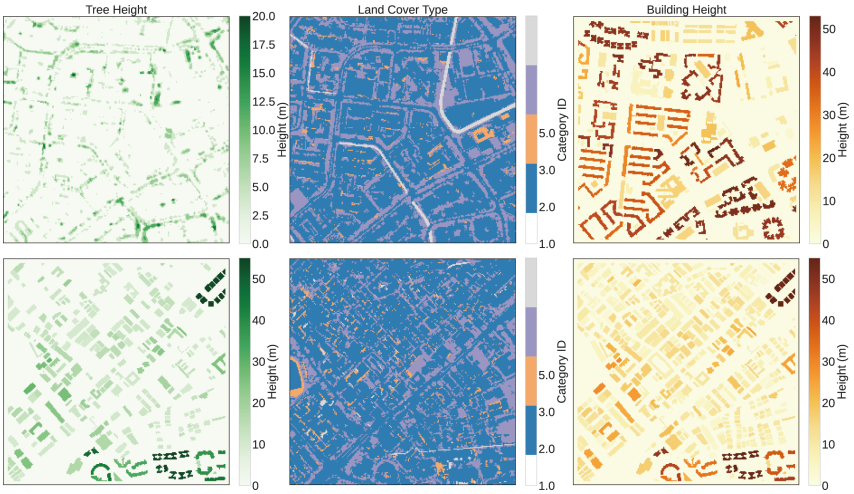

Figure A1: Visualization of the input data for two sample sites from the UMC4/12 dataset. The top row shows a dense, mixed-use urban area, while the bottom row depicts a residential area with more green space. Each column represents a different data layer: (left) Tree Height, (middle) Land Cover Type, and (right) Building Height.

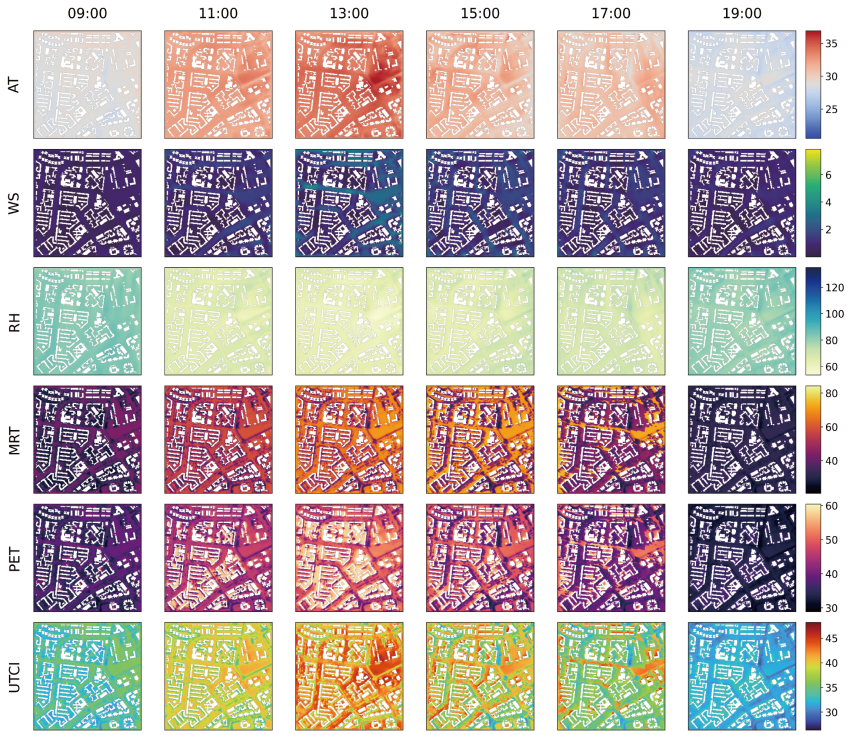

Figure A2: Visualization of the spatio-temporal simulation output for a single urban block. Each row corresponds to a different target variable: AT, WS, RH, MRT, PET, and UTCI. Each column represents a specific hour, showing the dynamic evolution of the microclimate from morning (09:00) to evening (19:00).

Table A2: Class definitions mapping land cover types to surface materials for ENVI-met simulation.

Type	Material	Class
Buildings	Pavement	1
Impervious surfaces	Pavement	1
Non-vegetated pervious surfaces	Terre battue	2
Vegetation with limited human management (w/ Tree Canopy)	Loamy Soil	3
Vegetation with limited human management (w/o Tree Canopy)	Loamy Soil	3
Vegetation with structure dominated by human management (w/ Canopy)	Loamy Soil	3
Vegetation with structure dominated by human management (w/o Canopy)	Loamy Soil	3
Freshwater swamp forest	Unsealed Soil	4
Freshwater marsh	Unsealed Soil	4
Mangrove	Deep Water	5
Water courses	Deep Water	5
Water bodies	Deep Water	5
Marine	Deep Water	5

Table A3: Material properties used in the ENVI-met model configuration.

Material	z_0 Roughness Length	Albedo	Emissivity
Pavement	0.010	0.3	0.9
Terre battue	0.010	0.4	0.9
Loamy Soil	0.015	0.0	0.9
Unsealed Soil	0.015	0.2	0.9
Deep Water	0.010	0.0	0.9

To expand the dataset while efficiently managing computational resources, we systematically partitioned the original 1 km² simulation results into 250m × 250m blocks with a 50-meter overlapping area. This resulted in a final dataset containing 396 unique urban blocks. Each block is discretized into 2,500 nodes. For each block, we provide a time series covering a 12-hour interval for 6 key microclimate and thermal comfort variables. Overall, the UMC4/12 dataset offers approximately 11.9 million high-quality spatio-temporal data points for each target variable, enabling the systematic evaluation of spatio-temporal prediction models in complex urban environments.

C Parameter Settings for Physics-Informed Graph Representation

This appendix details the key parameters used in the construction of the dynamic heterogeneous graph, as introduced in Section 4.1, and provides the rationale for their settings.

C.1 Parameter Rationale

Number of Nearest Neighbors (k). The value is chosen to balance informational richness with computational overhead. Allowing each node to connect to its eight most similar neighbors (consistent with the size of a Moore neighborhood) effectively captures non-local semantic information while avoiding the noise that could be introduced by connecting too many distant nodes. As demonstrated in the sensitivity analysis in Section E.2, this value represents the optimal trade-off between model performance and efficiency.

Maximum Shadow Extent (R_{max}^{shadow}). These upper limits are set to prevent unrealistically long shadows, which can occur at low solar elevation angles, from creating computational redundancy in the graph representation. The maximum shadow extent for buildings (15 grids, or 60m) is larger than that for trees (5 grids, or 20m), which is consistent with their typical differences in height and obstruction capacity in an urban environment.

Shadow Angle Width ($\Delta\phi_{shadow}$). This parameter expands the theoretical line-like shadow into an area of influence. This accounts for the apparent motion of the sun over an hour and the penumbra effect caused by diffuse light, making the shadow model more physically realistic.

Base Radius of Influence for Vegetation (R_{base}). The base radius of influence for vegetation is set to 5 grids (20m), based on the typical effective range of local cooling effects from single or small patches of green space reported in existing microclimate research (Kim et al., 2024).

Wind Effect Coefficient (λ_{wind}). As a modulation coefficient, a value of 0.3 is a relatively conservative choice. It allows the wind field to significantly guide the anisotropy of connections without completely dominating the graph structure, thus preserving the influence of other physical processes.

Maximum Reference Wind Speed (v_{max}). This value is used to normalize the actual wind speed. A value of 8.0 m/s was chosen as the reference upper limit as it represents the maximum wind speed historically observed in Singapore.

Table A4: Parameters for Dynamic Heterogeneous Adjacency Construction.

Parameter	Value	Description
Semantic Similarity Links		
k	8	The number of neighbors for semantic similarity links.
ϵ	1e-6	A small constant to avoid division by zero during feature normalization.
Shadow Links		
R_{max}^{shadow}	15	Maximum extent of building shadows (in number of grids).
R_{max}^{tree}	5	Maximum extent of tree shadows (in number of grids).
$\Delta\phi_{shadow}$	25.0°	The effective angular width for shadow calculations.
Vegetation Activity Links		
R_{base}	5	The base maximum radius of influence for vegetation activity (in number of grids).
Local Wind Field Links		
λ_{wind}	0.3	Coefficient that modulates the impact of wind direction on the connection range.
v_{max}	8.0 m/s	Used to normalize wind speed for calculating the wind modulation factor.

D Model Implementation Details and Hyperparameters

To ensure fairness, transparency, and reproducibility in our experimental comparisons, this appendix details the implementation specifics and key hyperparameter configurations for our proposed UrbanGraph model and all baseline models.

D.1 Baseline Models and Hyperparameter Settings

The following table summarizes the key hyperparameters for UrbanGraph and all baseline models used in the different experimental phases. In our comparative experiments, we strive to ensure a fair comparison by maintaining a similar model scale (i.e., hidden dimension size), such that performance differences primarily originate from the model architectures themselves.

D.2 Implementation Details for Cross-Paradigm Baselines

To compare our graph-based approach with traditional grid-based methods, we adapted the data input for certain baseline models.

Data Rasterization. For the CGAN-LSTM model, we convert the graph data at each timestep into a 50x50 grid image. Each node in the graph is mapped to a pixel in the image, where the pixel

Table A5: Key hyperparameters for the proposed model and all baseline models.

Model	Hidden Dim	Spatial Encoder	Temporal Encoder	Key Hyperparameters
UrbanGraph (Ours)	128/384*	RGCN(3)	LSTM(1)	`lr=0.001, batch_size=8, optimizer=Adam, weight_decay=1e-5`
GCN-LSTM	128	GCN(3)	LSTM(1)	same
GINE-LSTM	128	GINE(3)	LSTM(1)	same
RGCN-GRU	128	RGCN(3)	GRU(1)	same
RGCN-Transformer	128	RGCN(3)	Transformer	`d_model=128, nhead=4, num_encoder_layers=2`
CGAN-LSTM	128	U-Net	LSTM(1)	`lr_G=0.0002, lr_D=0.0002, beta1=0.5, lambda_L1=100`
GAE-LSTM	128	GAE(3)	LSTM(1)	`latent_dim=128, beta=0.1`
GGAN-LSTM	128	GGAN	LSTM(1)	`latent_dim=128, lr_G=0.0001, lr_D=0.0004, beta1=0.5`

**Note: The hidden dimension of UrbanGraph is 128 in the main model comparison phase. For the ablation and sensitivity analysis phases, it is set to 384 based on the results of Optuna optimization.*

value represents a key physical feature of the node (e.g., air temperature). The spatial relationships between nodes are implicitly represented by the adjacency of pixels on the 2D plane.

Model Implementation. We employ a classic U-Net as the generator for the CGAN and a PatchGAN as the discriminator. The model's task is to generate the prediction image for the next timestep based on a sequence of historical images. During training, we combine an L1 loss (with weight λ_{L1}) with an adversarial loss. The feature sequence extracted by the U-Net encoder is then fed into an LSTM module for temporal modeling.

E Detailed Analysis of Main Experiments

E.1 Detailed Ablation Studies

We conduct a series of ablation studies to systematically evaluate the contributions of the key components within the UrbanGraph framework.

Temporal Modeling. To validate the contribution of the Spatio-Temporal Evolution Module (LSTM), we compare the full Spatio-Temporal model against a variant where the LSTM module is removed. This variant performs independent predictions for each hour, thereby eliminating temporal dependencies. As shown in Figure A3 in the Appendix, the results demonstrate the effectiveness of temporal modeling. Our model's predictive accuracy (R^2) surpasses that of the variant across all prediction hours. Furthermore, Our model exhibits lower variance across multiple independent trials, indicating enhanced model stability.

Fusion Mechanism. We compare three different strategies for fusing the spatial node representations with the global dynamic features: Concatenation Fusion, Multiplicative Fusion, and Attention Fusion. As shown in Table A6, the simple concatenation strategy achieves the best performance across all evaluation metrics. This approach not only achieves the highest accuracy but also has the lowest computational load (FLOPs) and the fastest training and inference speeds.

Table A6: Comparison of different fusion strategies for combining spatial and global features. Best results are in **bold**.

Strategy	Flops	Test		Time Cost	
		Avg R^2	Avg RMSE	Training (epoch/s)	Inference/s
Attention	9.42×10^9	$0.8491 \pm .0052$	$1.0675 \pm .0254$	27.5568 ± 1.0310	$3.1741 \pm .1251$
Multiplicative	9.30×10^9	$0.8515 \pm .0020$	$1.0623 \pm .0317$	25.9573 ± 1.4409	$2.8543 \pm .2170$
Concatenation	$\mathbf{9.13 \times 10^9}$	$\mathbf{0.8542 \pm .0044}$	$\mathbf{1.0535 \pm .0338}$	$\mathbf{24.4823 \pm 0.9323}$	$\mathbf{2.6914 \pm .1404}$

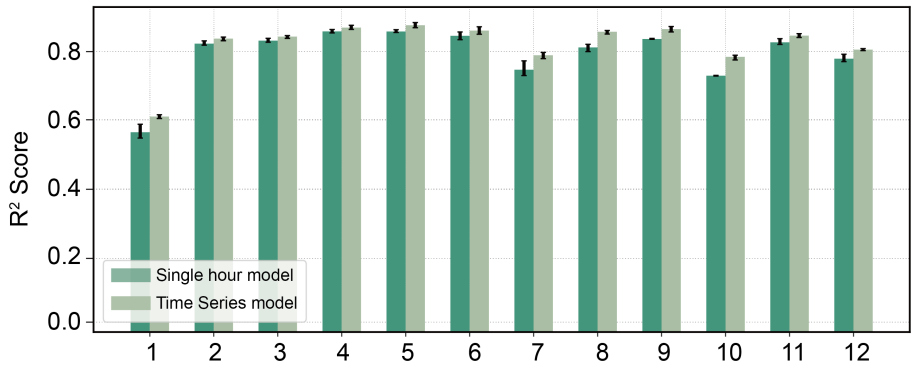

Figure A3: Hour-by-hour R^2 score comparison for the temporal modeling ablation study. The 'Time Series model' (our full UrbanGraph model) is compared against the 'Single hour model' (a variant without the LSTM module). The results demonstrate that explicitly modeling temporal dependencies leads to superior performance across the entire 12-hour prediction horizon. Error bars represent the standard deviation from multiple independent trials.

Warming-up Mechanism. We introduce a warming-up mechanism that initializes the LSTM's hidden state using the spatial features from the initial graph. This aims to provide the temporal prediction task with a starting point that is rich in physical priors. As shown in Table A7, removing this mechanism and using random initialization instead (the NP1 model) leads to a noticeable decline in performance, with the R^2 score dropping from 0.8629 to 0.8510.

Table A7: Effectiveness of the Warming-up Mechanism.

Model	R^2	MSE
Base	**0.8629**	**1.0976**
NP1	0.8510	1.1526

Prediction Head Architecture. We evaluate two strategies for multi-step prediction: a Single-Head architecture, which uses a single shared prediction head to generate predictions for all 12 hours at once from the final hidden state of the LSTM; and a Multi-Head architecture, which employs a separate prediction head for each future timestep. The results in Table A8 show that the single-head strategy performs better in terms of both predictive accuracy and computational efficiency (FLOPs). For a relatively short prediction horizon, the single-head architecture can more effectively leverage the final hidden state, which encodes information from the entire sequence, for joint prediction, thereby avoiding cumulative errors.

Table A8: Comparison between Single-Head and Multi-Head prediction architectures on the test set.

Strategy	Flops	Test		Time Cost	
		Avg R^2	Avg RMSE	Training (epoch/s)	Inference/s
Multi-Head	9.21×10^9	$0.8542 \pm .0044$	$1.0535 \pm .0338$	24.4823 ± 0.9323	$2.6914 \pm .1404$
Single-Head	$\mathbf{9.13 \times 10^9}$	$\mathbf{0.8603 \pm .0008}$	$\mathbf{1.0190 \pm .0421}$	$\mathbf{21.5903 \pm 3.1143}$	$\mathbf{2.4785 \pm .5019}$

Node Feature Augmentation. We compare the effects of using different node features as input. As shown in Table A9a, the model using aggregated neighbor features (Base) achieves the best performance. The model without any spatial information enhancement (M3) performs worse than the Base model. However, performance degrades when using only static topological features (such as degree centrality) or when combining them with aggregated neighbor features. This result suggests that introducing additional topological features in our task may add redundant information or noise, thereby impairing the model's predictive accuracy.

Input Feature Ablation. To verify the necessity of each input feature, we conduct a systematic ablation on the static node features $\boldsymbol{u}^{static}$, the temporal encoding features $\boldsymbol{u}_t^{time}$, and the global climate features $\boldsymbol{u}_t^{env}$. As shown in Table A9b, the baseline model (Base) that includes all three feature types performs the best. Removing any single feature type leads to a performance drop.

Notably, the model using only static node features (F1) shows the most significant degradation, with its R^2 score dropping from 0.8629 to 0.7179.

Table A9: Ablation studies for feature augmentation and input feature types.

(a) Data Augmentation.

Model	Neighbor	Structure	R^2	MSE
M1		√	0.8347	1.2798
M2	√	√	0.8462	1.2181
M3			0.8507	1.1696
Base	√		**0.8629**	**1.0976**

(b) Input Features.

Model	$\boldsymbol{u}^{static}$	$\boldsymbol{u}_t^{time}$	$\boldsymbol{u}_t^{env}$	R^2	MSE
F1	√			0.7179	2.0867
F2	√	√		0.8529	1.1423
F3	√		√	0.8519	1.1495
Base	√	√	√	**0.8629**	**1.0976**

Edge Types. To evaluate the specific contribution of each of the five proposed physics-informed and semantic edge types, we conduct an ablation study by systematically removing one edge type at a time. As shown in Table A10, the base model, which includes all five edge types, achieves the best performance. Removing any single edge type results in a decline in the model's predictive accuracy, demonstrating that both the physics-informed and semantic edges provide valuable inductive biases for the model. Notably, removing the *Local Wind* and *Shadow* edges leads to the most significant performance degradation, which underscores the importance of explicitly modeling time-varying physical processes. Furthermore, the performance drop caused by removing *Similarity* edges confirms the necessity of capturing non–local spatial interactions in urban microclimate prediction.

Table A10: Ablation study on different edge types. The checkmark (√) indicates that the corresponding edge type is included in the model.

Model	Tree activity	Similarity	Shadow	Local Wind	Internal	R^2	MSE
E1		√	√	√	√	0.8504	1.1534
E2	√		√	√	√	0.8531	1.1568
E3	√	√		√	√	0.8238	1.4960
E4	√	√	√		√	0.8155	1.4341
E5	√	√	√	√		0.8425	1.2403
Base	√	√	√	√	√	**0.8629**	**1.0976**

E.2 Sensitivity and Computation Performance Evaluation

Sensitivity to the Number of Neighbors (k). To investigate the model's sensitivity to the number of neighbors, k, used in constructing the Semantic Similarity Edges, we conducted tests with different values of k. As shown in Figure A4a, the model's performance (R^2) improves as k increases, reaching a peak at k=8 before exhibiting minor fluctuations. Considering that a larger k increases graph density and computational cost, we select the 'elbow point' of the performance curve, k=8, as the optimal configuration. The model is not highly sensitive to the choice of k within a certain range, demonstrating good robustness.

Sensitivity to Training Data Volume. To evaluate the model's data efficiency and generalization capability, we performed a sensitivity test on the amount of training data. We reserved a fixed 10% test set and incrementally increased the training set size using fractions of the remaining data, starting from 2%. As illustrated in Figure A4b, the results reveal a significant positive correlation between model performance and data volume, with all accuracy metrics improving substantially as the amount of data increases. However, the model also exhibits a clear diminishing returns effect: the majority of the performance gain occurs before the training data volume reaches 40-60%, after which the performance curve begins to plateau. Performance tends to saturate when approximately 90% of the available training data is used.

Computational Performance Evaluation. To assess the model's computational overhead in urban scenarios of varying complexity, we analyzed the relationship between the graph's structural properties (i.e., the number of nodes and edges) and computational costs (inference time and peak GPU memory usage). The analysis (Figure A5b,d) indicates that both inference time and memory consumption show a positive correlation with the number of edges in the graph (with R^2 values of 0.5976 and 0.4627, respectively). An interesting finding is that computational cost is negatively

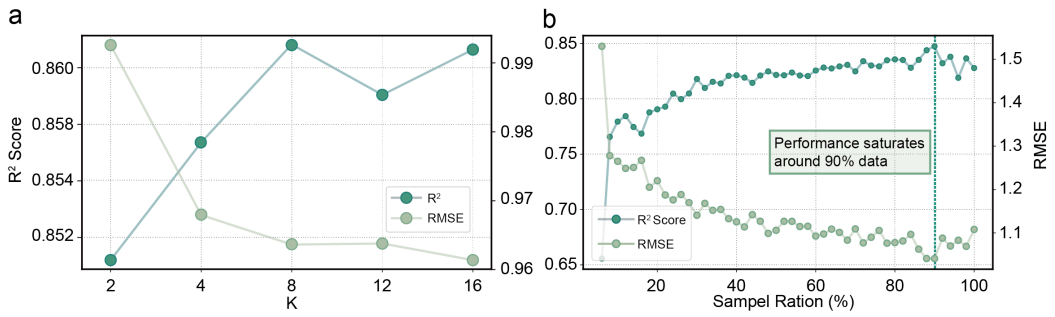

Figure A4: Sensitivity analysis of the model. (a) Model performance (R^2 and RMSE) on the test set versus the number of neighbors, k, for constructing similarity edges. The performance peaks at k=8. (b) Model performance as a function of the percentage of training data used. Performance gains show diminishing returns and begin to saturate at approximately 90% of the data.

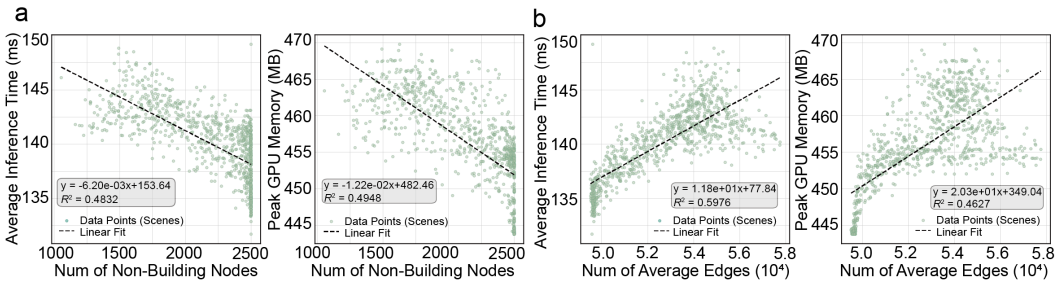

Figure A5: Computational performance analysis. (a) and (c) illustrate the negative correlation between inference time / peak GPU memory and the average number of non-building nodes per window. (b) and (d) show the positive linear correlation between computational costs and the average number of edges per window.

correlated with the number of non-building nodes (Figure A5a,c). This suggests that the number of non-building nodes can serve as an inverse indicator of a scenario's structural complexity: scenes with more open spaces (e.g., parks) typically have sparser graph structures and are therefore more computationally efficient. Furthermore, the linear relationship between cost and graph complexity suggests the feasibility of applying the model to larger areas.

E.3 Edge Features and Weights

To explore the potential of encoding richer physical information into the graph structure, we designed and evaluated an explicit scheme for edge attributes and edge weights in the early stages of our research. As mentioned in the main text, our final model did not adopt this design, as experimental results showed that introducing this explicit information did not lead to a significant performance improvement for the UTCI prediction task. This section details our initial exploratory design.

Edge Attribute Vector. In our initial design, each edge $e_{ij} \in \mathcal{E}_t$ in the graph carried a 5-dimensional attribute vector $\boldsymbol{a}_{ij} \in \mathbb{R}^5$ to encode rich spatio-temporal physical information. This vector was composed of the following components:

- **Euclidean Distance** $d(v_i, v_j)$: The straight-line distance between the nodes.
- **Relative Displacement** $(\Delta x, \Delta y)$: The difference in position in the grid coordinate system.
- **Wind Alignment** $(cos(\Delta\theta_{wind}))$: The cosine of the angle between the edge vector and the current hour's wind direction, used to quantify the convective influence of the wind.
- **Edge Type ID**: A categorical ID indicating which of the five relationship types the edge belongs to.

E.4 Edge Weight Calculation

To quantify the interaction strength between different nodes, we designed a dynamic scheme for calculating edge weights, w_{ij}. All edge weights start from a base value w_{base} (set to 1.0) and are dynamically modulated according to the following rule:

$$w_{ij} = w_{base}/(1 + \lambda \cdot d(v_i, v_j)/d_{grid}) \cdot \beta \cdot \gamma \tag{17}$$

The specific settings for each modulation factor are as follows:

- **Distance Decay (λ):** All edge weights are decayed based on their Euclidean distance. To better distinguish between non-local and local effects, we set a smaller distance decay factor, λ_{sim} (set to 0.005), for semantic similarity edges, while other edges based on physical proximity use a larger decay factor, λ_{phys} (set to 0.01).
- **Physical Process Enhancement (β):** Weights are further modulated by dynamic physical processes. For example, a shadow edge determined to be actively casting a shadow in the current hour has its weight multiplied by an enhancement factor, β_{shadow} (set to 1.2).
- **Source Node Attribute Influence (γ):** Weights are also influenced by the attributes of the source node. For instance, the weight of a vegetation activity edge is positively affected by the height of its source tree, h_{tree}, controlled by the modulation factor γ_{tree} (set to 0.2).

Although this scheme is theoretically more physically interpretable, our ablation experiments showed no improvement in predictive performance when introducing explicit edge attributes (E) and edge weights (EW) compared to a simpler model that only uses edge types (results shown in the table A11). This suggests that, for the UrbanGraph architecture and the UTCI prediction task, the model can effectively and implicitly learn the strength of these interactions from the dynamic graph topology and node features, without needing explicitly injected edge weights and attributes.

Table A11: Ablation study on explicit edge attributes (E) and edge weights (E_W).

Model	E	E_W	R^2	MSE
Base			**0.8629**	**1.0976**
EF	√		0.8530	1.1513
EFW	√	√	0.8586	1.2097

F Generalization and Visualization

F.1 Prediction results on different architectures

To provide a qualitative assessment of our model's performance, this section presents a visual comparison of the spatio-temporal prediction results between UrbanGraph and the four categories of baseline models. Each figure displays the ground truth, the predictions from UrbanGraph and representative baselines, and their respective prediction error maps (Prediction - Ground Truth) for a selected test scene at different hours of the day. White areas in the maps correspond to buildings, which are excluded from the analysis.

Table A12: Performance on other target variables.

Model	R^2
AT	$0.5650 \pm .1324$
WS	$0.7500 \pm .0176$
MRT	$0.8378 \pm .2005$
RH	$0.5159 \pm .2039$
PET	$0.8492 \pm .0517$

Figure A6 compares UrbanGraph with non-graph and homogeneous graph baselines, which represent fundamentally different approaches to spatial modeling.

Figure A7 provides a comparison against generative graph models and temporal variants, assessing different graph learning strategies and sequence modeling components.

F.2 Performance on multiple targets

To validate the robustness of our physics-informed graph representation and the UrbanGraph architecture, we trained the model using the same inputs for the remaining five target variables. The

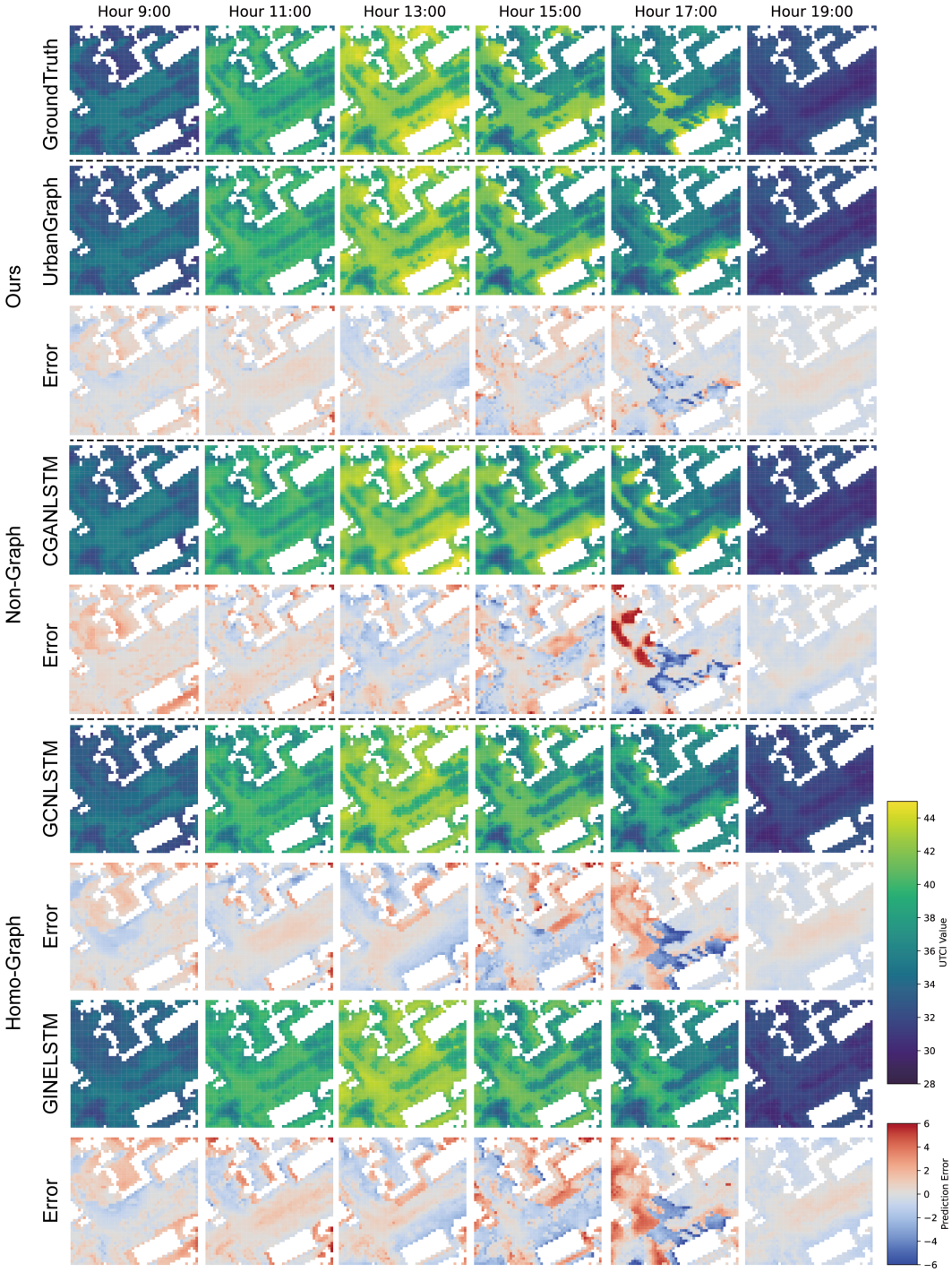

Figure A6: Visual comparison against Non-Graph (CGAN-LSTM) and Homogeneous Graph (GCN-LSTM, GINE-LSTM) baselines. Compared to the grid-based CGAN-LSTM, UrbanGraph better captures fine-grained spatial details. Unlike the homogeneous models that treat all interactions uniformly, UrbanGraph's heterogeneous approach leads to more physically consistent predictions and lower overall error.

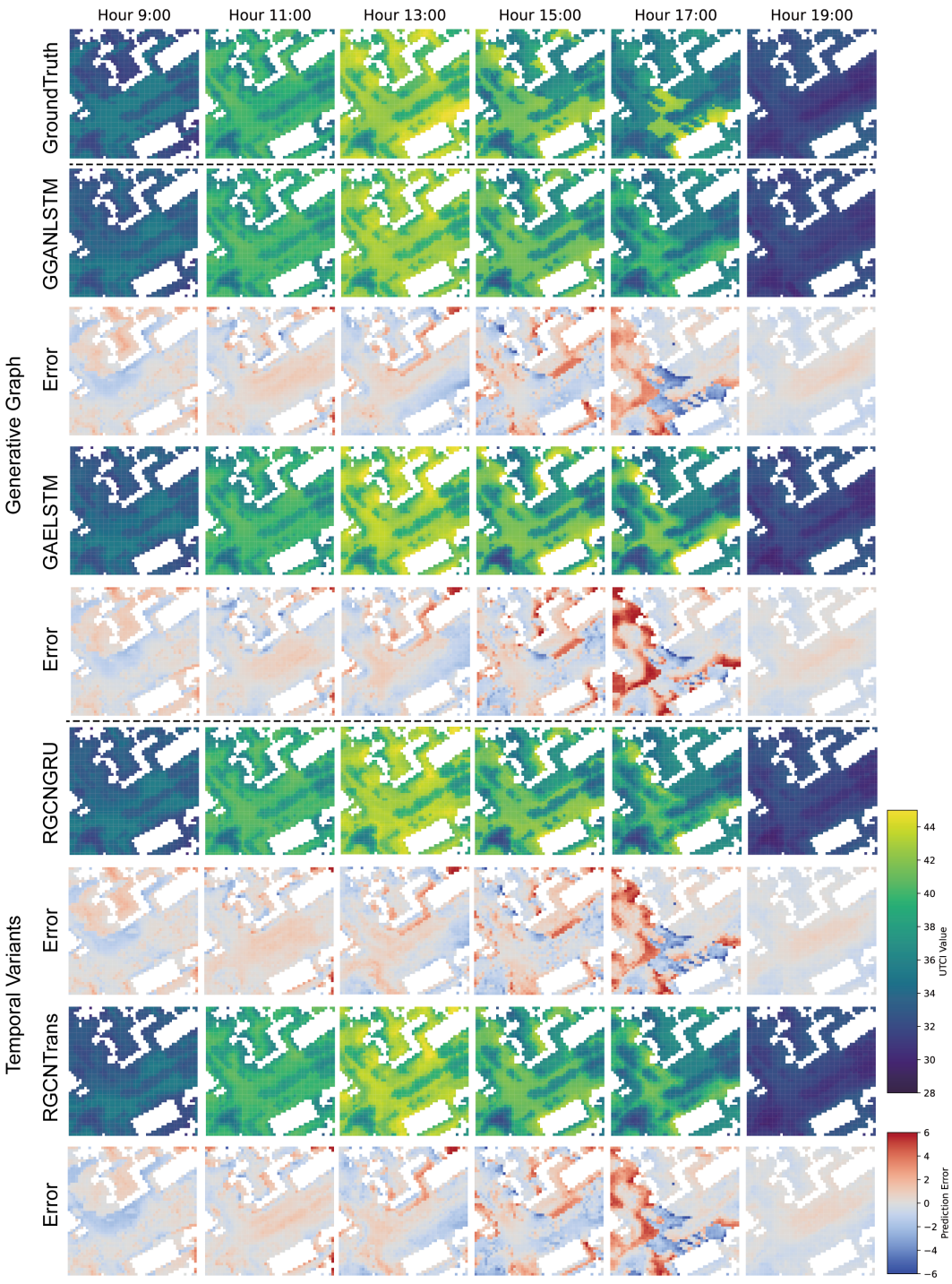

Figure A7: Visual comparison against Generative Graph (GGAN-LSTM, GAE-LSTM) and Temporal Variant (RGCN-GRU, RGCN-Transformer) baselines. UrbanGraph's physics-informed, deterministic graph construction (shown in Figure A6) avoids the higher errors seen in generative approaches. Furthermore, its LSTM component proves more effective at capturing long-term dependencies compared to the GRU and Transformer variants.

performance of the models for these five target variables is shown in Table A12. The R^2 scores for all models are above 0.5, with the performance on MRT and PET nearly matching the level achieved for UTCI, which demonstrates the effectiveness of our proposed method.

To provide a qualitative view of the model's generalization capabilities, the following figures visualize the spatio-temporal prediction results for the five other target variables.

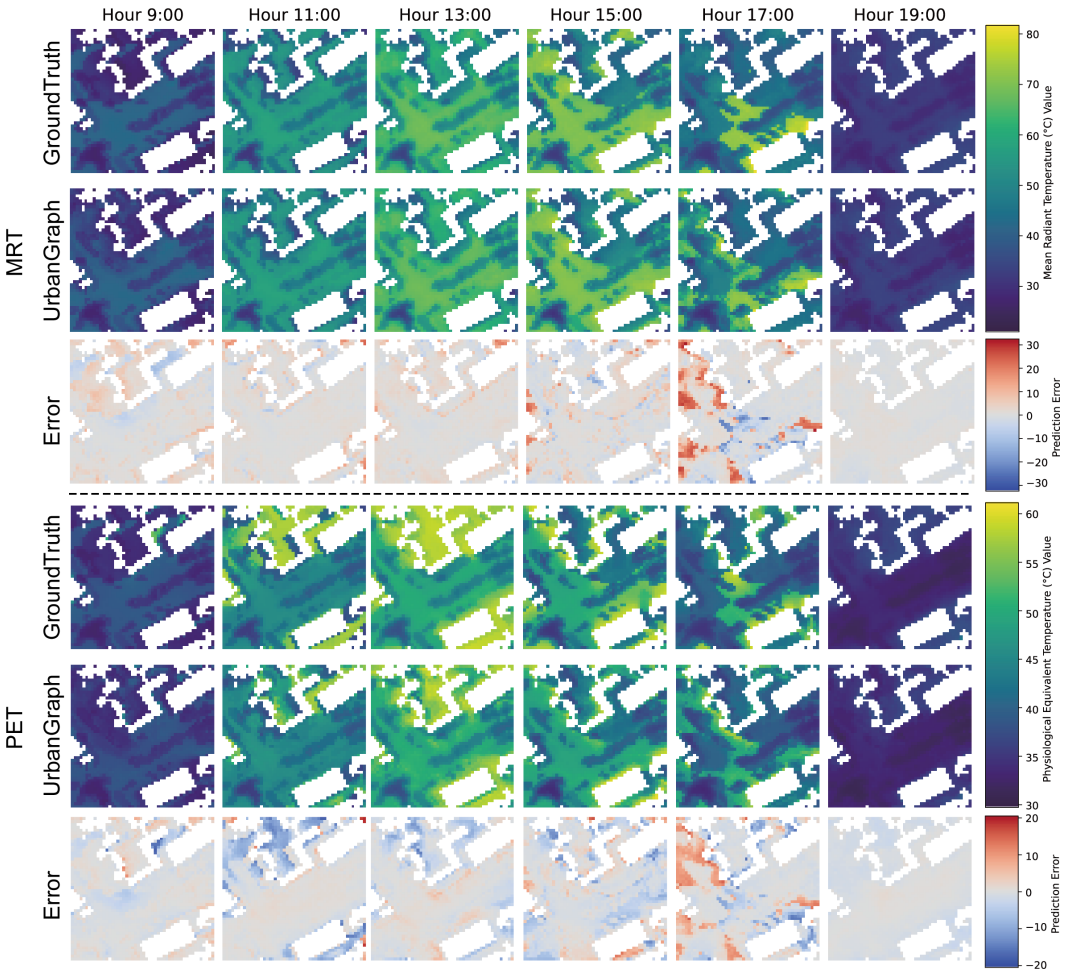

Figure A8: Qualitative prediction results for thermal comfort indices. This figure visualizes the performance of UrbanGraph on MRT and PET. Similar to the previous figure, each block compares the ground truth, model prediction, and the resulting error map, demonstrating the model's strong performance on composite indices.

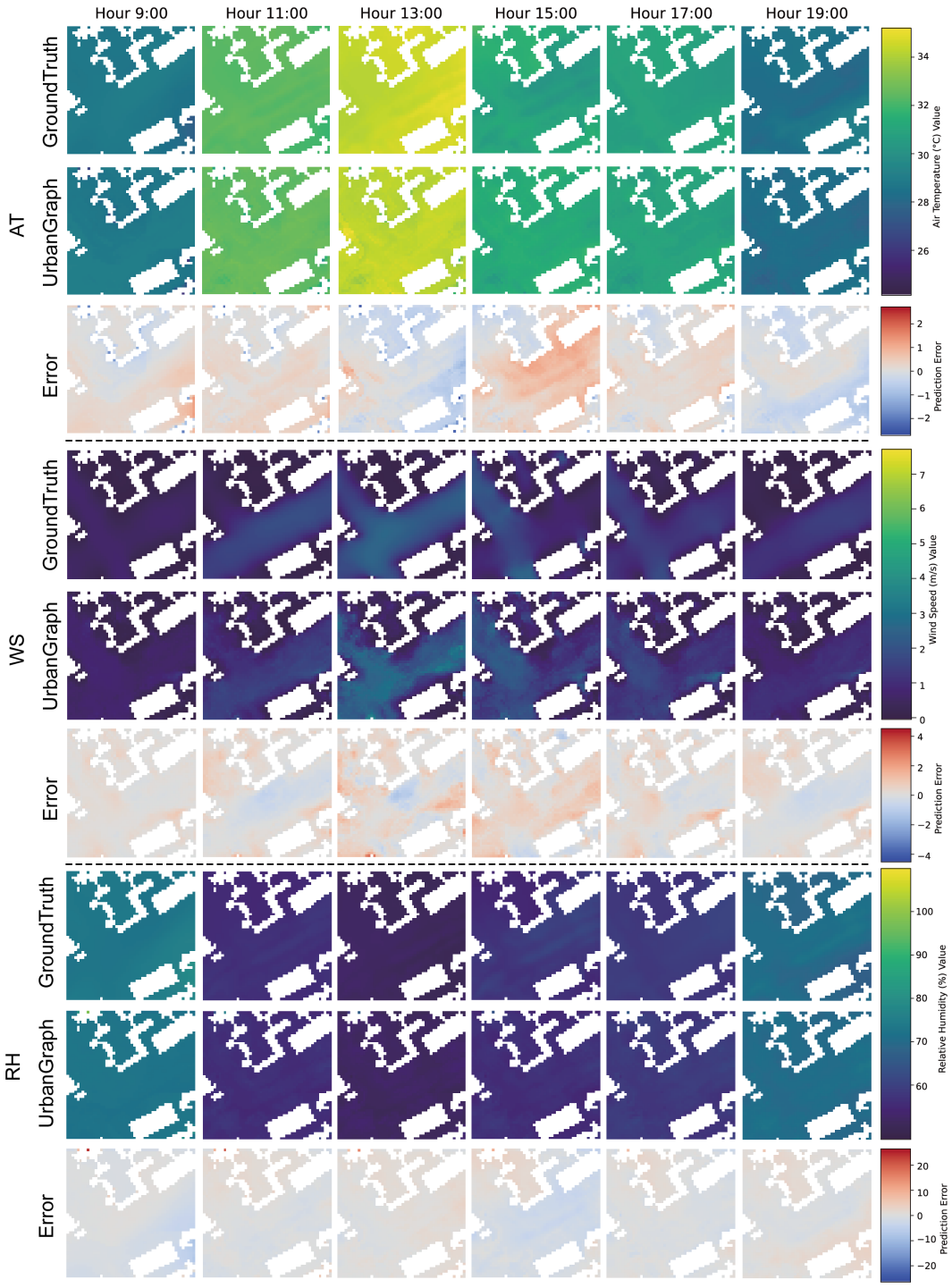

Figure A9: Qualitative prediction results for microclimate variables. This figure visualizes the performance of UrbanGraph on AT, WS, and RH. For each variable, the top row shows the ground truth, the middle row shows the model's prediction, and the bottom row displays the prediction error map across different hours.