Yin Yu[1], Peiyuan Li[2], Daning Huang[1], and Ashish Sharma[2,3,4]

[1]Department of Aerospace Engineering, Pennsylvania State University
[2]Discovery Partners Institute, University of Illinois System
[3]Department of Climate, Meteorology & Atmospheric Sciences, University of Illinois at Urbana-Champaign
[4]Environmental Science Division, Argonne National Laboratory

April 16, 2024

Abstract

**The manuscript has been submitted to Urban Climate for peer review.

Estimating street-level air temperature is a challenging task due to the highly heterogeneous urban surfaces, canyon-like street morphology, and the diverse physical processes in the built environment. Though pioneering studies have embarked on investigations via data-driven approaches, many questions remain to be answered. In this study, we leveraged an innovative framework and redefined the street-level temperature estimation problem using Graph Neural Networks (GNN) with spatial embedding techniques. The results showed that GNN models are more capable and consistent of estimating street-level temperature among tested locations, benefiting from its unique strength in handling extensive data over unstructured graph topology. In addition, we conducted an in-depth analysis of feature importance to enhance the model interpretability. Among the urban features analyzed in this study, the time-variant canopy density and meter-level land use data emerge as crucial factors. Our findings highlight GNN's high potential in capturing the complex dynamics between urban elements and their impacts on microclimate, thus offering valuable insights for comprehensive urban data collection and urban climate modeling in general. Collectively, this study also contributes to urban planning and policy by providing avenues to enhance city resilience against climate change, thereby advancing the agenda for environmental stewardship and urban sustainability.

The manuscript has been submitted to Urban Climate for peer review.

Street-level temperature estimation using Graph Neural Networks: Performance, feature embedding and interpretability

Yin Yu[1], Peiyuan Li[2], Daning Huang[1], and Ashish Sharma[2,3,4]

[1]Department of Aerospace Engineering, Pennsylvania State University, University Park, PA, 16802, USA

[2]Discovery Partners Institute, University of Illinois System, Chicago, IL, 60606, USA

[3]Department of Climate, Meteorology & Atmospheric Sciences, University of Illinois at Urbana-Champaign, Champaign, IL, 61820, USA

[4]Environmental Science Division, Argonne National Laboratory, Lemont, IL, 60439, USA

March 25, 2024

Abstract

Estimating street-level air temperature is a challenging task due to the highly heterogeneous urban surfaces, canyon-like street morphology, and the diverse physical processes in the built environment. Though pioneering studies have embarked on investigations via data-driven approaches, many questions remain to be answered. In this study, we leveraged an innovative framework and redefined the street-level temperature estimation problem using Graph Neural Networks (GNN) with spatial embedding techniques. The results showed that GNN models are more capable and consistent of estimating street-level temperature among tested locations, benefiting from its unique strength in handling extensive data over unstructured graph topology. In addition, we conducted in-depth analysis of feature importance to enhance the model interpretability. Among the urban features analyzed in this study, the time-variant canopy density and meter-level land use data emerge as crucial factors. Our findings highlight GNN's high potential in capturing the complex dynamics between urban elements and their impacts on microclimate, thus offering valuable insights for comprehensive urban data collection and urban climate modeling in general. Collectively, this study also contributes to urban planning and policy by providing avenues to enhance city resilience against climate change, thereby advancing the agenda for environmental stewardship and urban sustainability.

Keywords

Street-level temperature; Graph Neural Networks; Spatial embedding; Urban climate informatics; Urban features

1 Introduction

The increased population in global cities has led to fast and extensive urban expansion and densification in the recent decades (United Nations, 2019). Urban dwellers are believed to be more susceptible to environmental hazards, especially extreme weather events with amplified frequency, intensity, and duration by global climate change (Myhre et al., 2019; Perkins-Kirkpatrick et al., 2020). These events affect urban areas disproportionally, depending on the geographical, morphological, and thermodynamic features (Oke, 2008; Oke et al., 2017). Recent years have witnessed numerous studies working to improve the accuracy and spatial resolution of urban environmental modeling, aiming to address the challenges in quantifying the drastic inter-urban and intra-urban variabilities led by the highly heterogeneous built environment (Scott et al., 2017; Kousis et al., 2021; Cao et al., 2022).

There are two major barriers for the current process-based urban climate models to achieve ideal performance: (1) the lack of accurate data from the real world for precise parameterization; and (2) the lack of physical representations on certain processes. Accordingly, the on-going effort in urban climate research community diverts into two mainstreams. One direction focused on a more representative and realistic description of urban fabric, exemplified by the local climate zone (LCZ) classification scheme (Demuzere et al., 2021; Stewart et al., 2012; Kim et al., 2021), WUDAPT (Ching et al., 2018), and other urban canopy parameter databases (Hammerberg et al., 2018; Pilant et al., 2020; B. Chen et al., 2021) with exceptional spatial resolutions ranging from 100m to 1m. The other attempts to improve model performance by including detailed parameterizations, such as the inclusion of building energy exchange (Kondo et al., 2005; Jin et al., 2021), tree shading (Krayenhoff et al., 2020; C. Wang et al., 2021), ecohydrological processes (Stavropulos-Laffaille et al., 2018; Meili et al., 2020), and physiological functions (P. Li and Z.-H. Wang, 2020) in urban canopy models. These models can resolve up to a few hundred meters, but are more commonly seen at 1km resolution. The integration of these two streams, as represented by Meyer et al. (2020) and Ribeiro et al. (2021), has demonstrated enhanced performance over the less sophisticated process-based models, offering valuable insights on the in-canyon microclimate dynamics. Nevertheless, the improvements sometimes can be disproportional to the increased burden on computational cost, leading to a "resolution-coverage dilemma". Practically, it is nearly impossible for process-based urban climate models to achieve city-wide simulations with meter-level resolution in a near-real-time manner.

To address this challenge, some pioneering studies have investigated the data-driven approach by leveraging the state-of-the-art machine learning (ML) technology and the contemporary advancements in urban climate informatics (Middel et al., 2022; P. Li and A. Sharma, 2024b). Recent research, aided by the high-precision remote sensing (D. Yu et al., 2023), distributed sensor network (Catlett et al., 2017; Y.-C. Chen et al., 2019), and mobile measurement (A. Wang et al., 2023), has yielded promising results in estimating land surface and air temperatures (Venter et al., 2020; S. Sharma et al., 2023), air quality (Gitahi et al., 2020; Guo et al., 2022; A. Wang et al., 2023), and flooding conditions (Silverman et al., 2022; Tien et al., 2023) with exceptional spatial granularity, down to 10 meters. These studies provide insights into the actual environmental conditions experienced by urban residents, thus holding profound implications for research on walkability, heat-related mortality, hazard exposure, and environmental/climate justice. Moreover, they can guide meaningful real-world mitigation and adaptation efforts while enhancing our understanding of general hydroclimate dynamics in complex urban environments. One major gap, nevertheless, is that observation-based approaches usually lack forecasting capabilities, as they require data as a priori condition for the subsequent estimations. The availability of remote sensing imagery can be constraint by cloud cover. Weather conditions also create operational barriers for mobile measurements.

More recently, P. Li and A. Sharma (2024b) introduced a novel hybrid ML framework that integrates a meso-scale weather forecast model, detailed urban geographical datasets, and a set of street-level sensors to estimate in-canyon air temperature. This innovative endeavor not only grants predictive capabilities, but also provides point-scale temperature estimations that surpass conventional notions of spatial resolution, enabling the users to analyze thermal environment at specific locations using either historical hindcast data, near-real-time weather forecasts, or future climate projections. The inclusion of regional scale weather conditions in this hybrid approach also empowers the ML model with knowledge of synoptic weather dynamics, therefore producing more trustworthy estimations. Nevertheless, pivotal inquiries persist concerning the sensitivity and interpretability of such data-driven models. Specifically, there is a pressing need to investigate the significance of the urban features to street-level air temperature. Further studies on model sensitivity are also anticipated to test the robustness of the framework and enhance our comprehension of the hybrid approach.

In this study, our goal is to further advance the method presented in P. Li and A. Sharma (2024b) by introducing a more sophisticated ML algorithm, Graph Neural Networks (GNN), to the hybrid modeling framework. GNN is a recent variant of deep learning algorithms and has a specialty in the modeling of unstructured data defined on graphs or networks (Scarselli et al., 2009). Its applications to climate science have covered a wide range of topics, including the predictions of global weather (Keisler, 2022; Lam et al., 2023), regional heatwaves (P. Li, Y. Yu, et al., 2023), air quality (S. Wang et al., 2020; Ejurothu et al., 2023; Ma et al., 2023), frost (Lira et al., 2022), and precipitation (Y. Chen et al., 2024), which demonstrates a high potential to tackle the complex urban environment with extensive geospatial datasets. Another merit of GNN specific to the street-level downscaling problem is its architectural advantage. Since street-level sensors can only provide ground truth at distributed locations, this characteristic makes this street-level downscaling challenge differs fundamentally from the downscaling of climate simulations of two spatial continuous layers with different resolutions. The latter question has been widely addressed using Generative adversarial networks (GANs), Convolutional Neural Networks (CNNs), and other super-resolution algorithms, with examples highlighted in research by F. Wang et al., 2021; J. Wang, Liu, et al., 2021; Singh et al., 2023, respectively. These methods, primarily optimized for processing images characterized by inherent smoothness and continuity, thus do not directly apply to the discrete nature of the downscaling task discussed herein. In contrast, GNN can adeptly handles both discrete and continuous datasets by organizing data into a graph structure. In addition, when compared to conventional algorithms like Random Forest (RF), Gaussian Process Regression (GPR), XGBoost, and Support Vector Machine (SVM), which process temporal dynamics independently at each node, the structural advantage of GNN can facilitate the dynamic information exchange between nodes through their connecting edges. Therefore, GNN emerges as a tailored solution to address the distinct challenges highlighted in this research.

In addition to model development, we adopt GNNExplainer (Ying et al., 2019), a post-hoc algorithm, to examine the reliance of the predicting mechanisms on certain model inputs, aiming to enhance the interpretability of the trained GNN models and improve the general understanding of urban microclimate dynamics from an ML persepctive. Collectively, our investigations will contribute from four aspects: (1) to frame a hyper-local downscaling problem into GNN architecture, thus facilitating the implementation of advanced ML algorithms in urban climate modeling to overcome the limitations inherited from conventional modeling methods; (2) to harness the existing data inventory and improve the hyper-local temperature estimation; (3) to quantify the importance of urban climate informatics, thereby precisely guiding future observation and data curation endeavors; and (4) to test and validate the feasibility of GNN in the hybrid modeling framework presented in (P. Li and A. Sharma, 2024b). The findings will shed light on the evolution of urban climate informatics and have the potential to revolutionize urban land surface modeling, thus paving the way for more accurate and resilient urban planning and management strategies.

The following manuscript is organized into 5 sections, with Section 2 providing detailed descriptions of the urban datasets used in this study, followed by Section 3, digesting how these datasets are integrated for the street-level temperature downscaling problem. Section 4 elucidates the modeling methods of GNN and GNNExplainer, including model architecture, configuration, and evaluating metrics. Modeling results and discussions can be found in Section 5, followed by concluding remarks in Section 6.

2 Data Preparation

We identify three data components that are essential for addressing the urban downscaling challenge and for the efficacious deployment of the ML model (P. Li and A. Sharma, 2024b):

1. **Temporal dynamics layer:** A low-resolution dataset to encapsulate the temporal dynamics of the system, providing the synoptic view of the meteorological conditions over time.

2. **High-resolution ground truth data:** This dataset serves as the target for ML model training, comprising precise temperature recordings from an extensive observation network that anchors both the ground truth and the learning objective.

3. **Geographical feature set:** A collection of urban attributes crucial for enabling the ML model to understand the spatiotemporal interplay between low- and high-resolution datasets, thereby capturing the nuanced microclimatic variations within urban landscapes.

For the first component, we employ the weather hindcasts, offering a comprehensive perspective of meteorological conditions and approximate surface weather across Chicago. The second component comprises precise temperature measurements from a comprehensive observation network, acting as both the ground truth and the learning objective. For the third component, we identified and extracted various urban features that have a significant impact on the microscale climate within urban settings. Further details on each component are elaborated upon in subsequent sections. For clarity, a comprehensive table summarizing all variables and features utilized in this study is provided in Table 1.

2.1 Weather hindcasts

In this study, we use the Weather Research and Forecast (WRF) model version 4.0 (F. Chen, Kusaka, et al., 2011; Skamarock et al., 2021) to reconstruct the near-surface meteorological conditions at 1km spatial resolution and hourly intervals, serving as the low-resolution layer of this downscaling problem. WRF is a fully compressible, Euler nonhydrostatic Continuous weather prediction and atmospheric simulation system designed for both atmospheric research and operational forecasting applications (Skamarock et al., 2021) that has been widely adopted in numerous regional and global atmospheric and meteorological studies. Specific to this study, we set up three two-way nested domains with the outermost boundary covering the east-north central region of the Midwest US and the innermost domain covering the City of Chicago and its surrounding metropolis. The spatial resolutions of the three domains are 9 km, 3 km, and 1 km, respectively. The lateral boundary conditions are from North American Regional Reanalysis (NARR) from the National Center for Environmental Prediction (NCEP, https://rda.ucar.edu/datasets/ds608.0/) with a 32-km horizontal spatial resolution and a 3-hr temporal resolution. We adopt the single-layer urban canopy model for impervious urban surfaces (F. Chen, Kusaka, et al., 2011) and Noah-land surface model (F. Chen and Dudhia, 2001) for natural land and the previous portion of the urban grids. We also use WRF Single-Moment 6-class microphysics scheme, which is suitable for high-resolution simulations (Hong and J.-O.J., 2006). Longwave and shortwave radiation is parameterized using the Rapid Radiative Transfer Model (Iacono et al., 2008). Sub-grid scale cumulus convective parameterization is turned on only for the two outermost domains (9km and 3km) corresponding to the Kain-Fritsch scheme (Kain, 2004). The planetary boundary layer is simulated by Yonsei University scheme (Hong, Noh, et al., 2006), while the surface layer is parameterized by Monin-Obukhov similarity scheme. The configuration and physical schemes were well tested in multiple previous studies over Chicago (A. Sharma et al., 2017; P. Li, A. Sharma, et al., 2023).

The hindcast covers two summers in 2018 and 2019 (May 1st to Aug 31st, 123 days). We select six variables from WRF, namely air temperature and humidity 2 meters above the ground, land surface temperature, soil temperature, downwelling shortwave radiation, and wind speed 10 meters above the ground, as the input of the subsequent ML model (WH variables in Table 1). These variables were validated against the observations from ground weather stations from National Center of Environmental Information (NCEI) to ensure WRF captured the synoptic weather dynamics. It is worth noting that despite we did not calibrate the parameters or physical schemes in WRF model, the simulation result is acceptable with an RMSE of 2.5 ^oC for daily mean air temperature, which is widely accepted among existing urban climate modeling studies.

2.2 Temperature observation network

The Array of Things (AoT) project started in 2018 and was designed to monitor the urban environment of Chicago via a dense observational network (Catlett et al., 2017). The measurement sensors contain an array of environmental sensors that are mounted on existing urban infrastructures (such as traffic light poles, building walls, bus stations, etc.) at over 100 locations in Chicago city. The sensors measure the meteorological variables, air quality, noise level, and traffic at sub-minute intervals. These measurements are wirelessly transmitted to a data center in a real-time manner and are compiled into a complete dataset for public access. Most of the sensors are located 2 to 4 meters above the ground thus representing street-level conditions reasonably well.

During the designed operation period (Jan 2018 to Sept 2021), the AoT network effectively collected air temperature readings from 106 sensors during 2018 and 2019. We carefully calibrated the temperature recording from AoT using the nearby research-grade weather stations (P. Li and A. Sharma, 2024b) to ensure the data quality of these low-cost sensors. But due to their low-cost nature, calibrated temperature readings from AoT sensors may still associated with uncertainties, bias, and errors. Nevertheless, we treat the AoT observations as the best proxy for the "ground truth" of the urban environment given the current data scarcity in the urban environment. The screened dataset contains continuous timeseries measurement of air temperature over 53 locations, and 15 of them have both measurements over summers of 2018 and 2019. This leads to an equivalent of 200,736 measurement hours as the total data points used in GNN development. The complete set of AoT data can be downloaded with additional information at http://arrayofthings.github.io/.

2.3 Detailed urban features

The detailed urban morphological and geographical features are derived from a suite of high-resolution urban-orientated datasets over Chicago. These include two sets of land use classifications, impervious surface fractions, vegetated surface fractions, tree canopy coverage, building height, and tree height. We derive these geospatial information from independent data sources.

The 2018 parcel-level land use inventory (LUI) for the City of Chicago can be found at the Chicago Metropolitan Agency for Planning (CMAP) in vector format (https://www.cmap.illinois.gov/data/land-use/inventory). The LUI classifies the land use into 10 major and 56 minor categories. We convert the vectorized shapefile into a raster layer with 1-meter resolution to align the spatial resolution of the other datasets. In addition, we also adopt the land cover types from the National Land Cover Database (NLCD) with 30-meter resolution. Compared to NLCD, the parcel-level LUI has a more detailed classification based on the primary use of the urban land, but NLCD provides the development intensity as additional information on the urban features.

The tree canopy coverage, impervious and vegetated fractions are derived from The Meter-scale Urban Land Cover (MULC) from the US Environmental Protection Agency (EPA). This urban-oriented dataset has a good representation of the urban landscapes with exceptional resolution and accuracy (Pilant et al., 2020). It classifies urban land into 10 categories, including impervious surfaces, trees, shrubs, grass, water, crops, etc. These classifications are converted to binary maps indicating the spatial distribution of different land cover. It is noteworthy that impervious surfaces in MULC consist of roads and buildings. The distinction between them needs to rely on the additional height information. We adopt HeIght map of Tree And Buildings in Chicago (HiTAB-Chicago) for an accurate 3-dimensional description of the urban morphology. HiTAB-Chicago is a LiDAR-based digital elevation models with a 1-meter resolution containing tree and building heights as separate layers (P. Li and A. Sharma, 2024a). Unlike the categorical or binary classifications, the height information is in continuous values, thus enriching the data types in this regression task.

To align the data format and resolution of these seven geospatial maps (CMAP, NLCD, three from

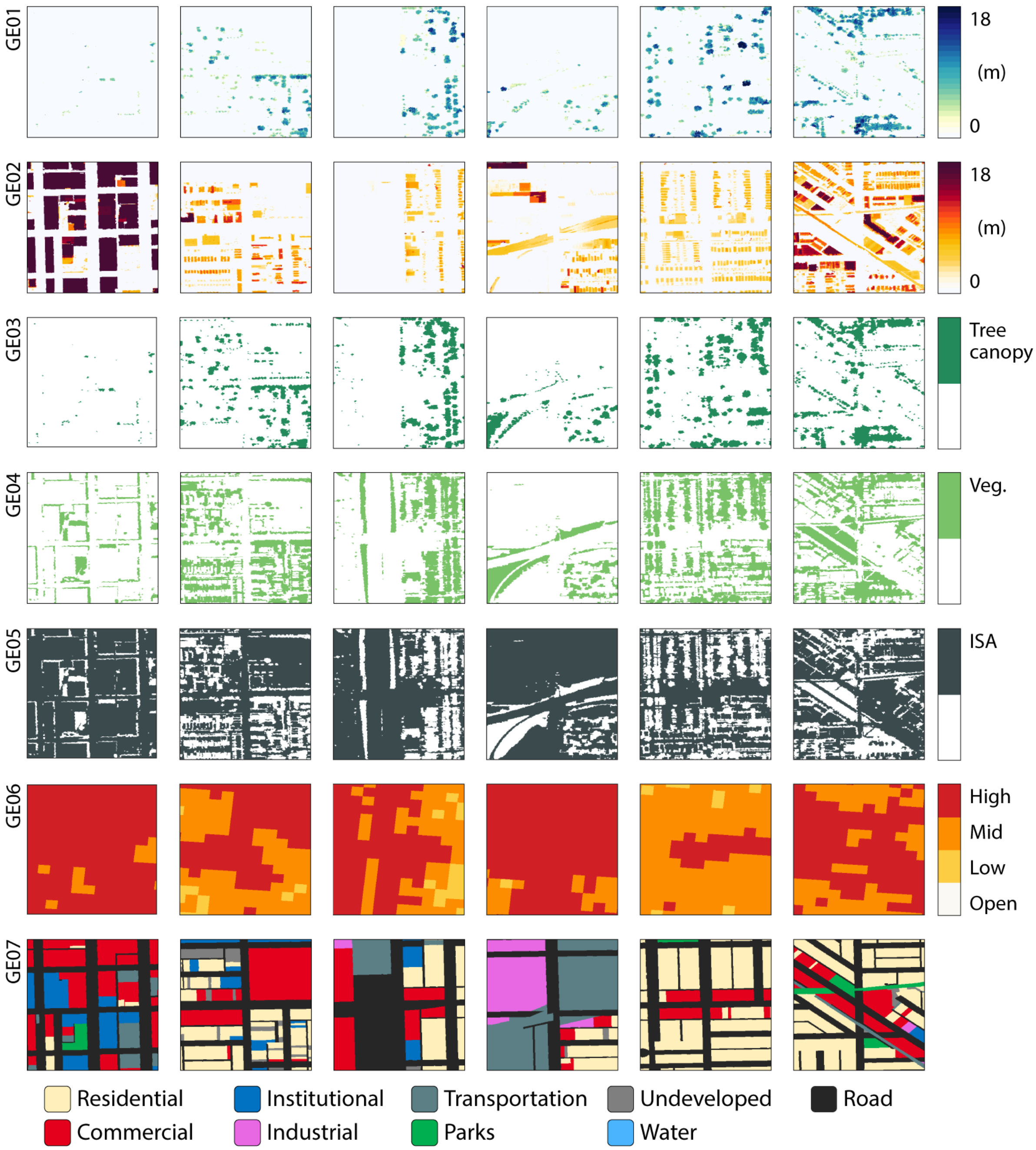

Figure 1: Embedded geofeatures at six exemplary sensor locations with a patch size of 400m by 400m. GE01 - Canopy height; GE02 - Building height; GE03 - Canopy patch; GE04 - Vegetated patch; GE05 - Impervious surface area patch; GE06 - NLCD development intensity; GE07 - CMAP land use classification.

MULC, two from HiTAB), we convert them into raster layers with 1-meter resolution. Subsequently, the landscape patch is extracted as a 400m by 400m grid centered by AoT sensors (Fig. 1). These patches will be aggregated and embedded as the inputs of GNN model (see Section 3.2).

In addition to the variables used as spatial embedding, we also include the statistical moments (i.e., averages and standard deviations) as model input. These include the fractions of impervious, vegetation, water, tree canopy, and different development intensities within the 400m by 400m grid, as well as the mean, maximum, and standard deviation of the tree and building heights. We also include the mean height of trees and buildings south of the observation sensors to better reflect the shading effect for cities in the Northern Hemisphere. Due to its special geographic location, the urban environment in Chicago is under the influence of the lake breeze effect (J. Wang, Qian, et al., 2023). Therefore, the distance to Lake Michigan is added as an attribute for each AoT sensor.

It is noteworthy that the urban features mentioned above do not change over time. But plant leaf density will change gradually due to phenology during their growing period in the summer months. To inform the model with this variation, we extract the timeseries of canopy coverage, leaf area index (LAI), and normalized difference vegetation index (NDVI) from 10-day 300-meter Copernicus Global Land Service (CGLS) products. Nevertheless, their spatial resolutions are relatively low compared to the other datasets. We use these indices as independent information on the temporal scale.

3 Problem Statement

The urban downscaling problem aims to refine coarse-grained meteorological data into high-resolution, street-level temperature predictions across urban landscapes. The core objective is to accurately predict an array of street-level temperatures, denoted as $\mathbf{T}_a$, at different sensor locations within an urban area, e.g. the AoT network. This process leverages a combination of geospatial characteristics and sensor data measurements. Here, we re-construct the AoT network as a graph and carefully craft geospatial attributes as feature vectors (hereinafter referred to as geofeatures) over each sensor. The subsequent section discusses the details of graph representation of the AoT network and is succeeded by a discussion on feature selection, which describes how a comprehensive dataset is compiled into informative inputs for the modeling process.

3.1 Graph representation

A graph or network represents data through a set of nodes, a set of edges that defines the pairwise relations of the corresponding nodes. We conceptualize the AoT network as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, which contains the N measurement sensors as the nodes $\mathcal{V}$, the edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ as the connections between each pair of

Table 1: List of all features that are used as model inputs.

Notation	Variable	Type	Source
	Weather hindcast (WH)		
WH01	2-meter air temperature	Continuous	WRF
WH02	2-meter air humidity	Continuous	WRF
WH03	Soil temperature	Continuous	WRF
WH04	Surface temperature	Continuous	WRF
WH05	Solar irradiance	Continuous	WRF
WH06	Wind speed	Continuous	WRF
	Auxilary (AX)		
AX01	Hour of the day (UTC)	Continuous	
AX02	Month of the observation	Continuous	
AX03	Year of observation	Continuous	
	Patch-embedded urban features (GE)		
GE01	Height - Tree	Continuous	HiTAB - Chicago
GE02	Height - Building	Continuous	HiTAB - Chicago
GE03	Patch - Tree	Binary	EPA - MULC
GE04	Patch - Vegatation	Binary	EPA - MULC
GE05	Patch - Impervious	Binary	EPA - MULC
GE06	Classification - NLCD	Category	NLCD
GE07	Classification - CMAP	Category	CMAP
	Averaged urban features (GA)		
GA01	Mean height - Tree	Continuous	HiTAB - Chicago
GA02	Max height - Tree	Continuous	HiTAB - Chicago
GA03	Height std. - Tree	Continuous	HiTAB - Chicago
GA04	Mean height - Tree in south	Continuous	HiTAB - Chicago
GA05	Timeseries - Vegetation coverage	Continuous	CGLS
GA06	Timeseries - LAI	Continuous	CGLS
GA07	Timeseries - NDVI	Continuous	CGLS
GA08	Mean height - Building	Continuous	HiTAB - Chicago
GA09	Max height - Building	Continuous	HiTAB - Chicago
GA10	Height std. - Building	Continuous	HiTAB - Chicago
GA11	Mean height - Building in south	Continuous	HiTAB - Chicago
GA12	Fraction - High dev. intensity	Continuous	NLCD
GA13	Fraction - Medium dev. intensity	Continuous	NLCD
GA14	Fraction - Low dev. intensity	Continuous	NLCD
GA15	Fraction - Open development	Continuous	NLCD
GA16	Fraction - Impervious surface	Continuous	EPA - MULC
GA17	Fraction - Low vegetation	Continuous	EPA - MULC
GA18	Fraction - Water	Continuous	EPA - MULC
GA19	Fraction - Tree canopy	Continuous	EPA - MULC
GA20	Distance to lake Michigan	Continuous	NLCD

nodes, and the edge weights $\mathcal{W}$ quantify the correlation between the states of two connected sensors. Every measurement sensor that is connected to a node of interest (NoI) by an edge is known as a neighbor of the NoI. The graph for the AoT network is symmetric, meaning that if $(i,j) \in \mathcal{E}$ then $(j,i) \in \mathcal{E}$. The strength of connection between two connected nodes i and j, or edge weight w_{ij}, is determined by a combination of

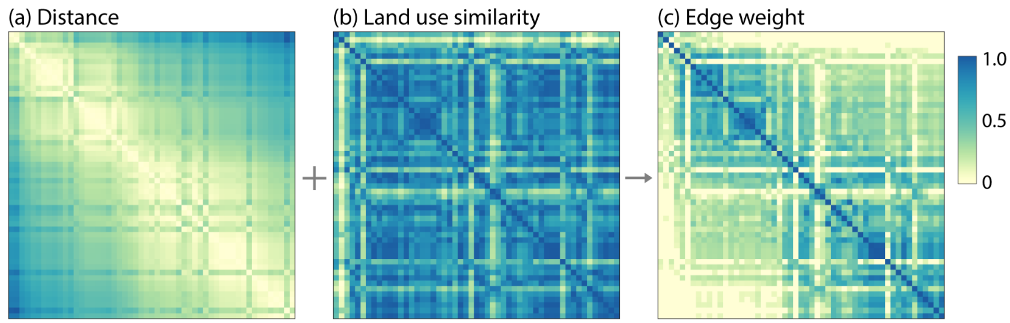

Figure 2: Matrices of normalized (a) node distance, d_{ij}; (b) land use similarity, s_{ij}; and (c) the resulted edge weight, w_{ij}, derived from the AoT observation network after quality control.

the land use similarity (s_{ij}) and physical distance between them (d_{ij}), as expressed by

$$w_{ij} = |s_{ij} \exp(-d_{ij})| \tag{1}$$

where s_{ij} = is calculated as the correlation between the vectors of the land use fractions over the paired nodes, expressed as $s_{ij} = \mathrm{Cor}(F_{i,\mathrm{LUI}}, F_{j,\mathrm{LUI}})$. This formulation ensures that sensors which are both highly similar in terms of land use patterns and proximate in physical distance exhibit a stronger linkage within the graph. Edges bearing weights below a threshold α are disregarded to maintain graph sparsity, enhances computational efficiency, model scalability, and focuses on the most significant relationships among the sensors. We further define the adjacency matrix $\mathbf{A}$, which encapsulates the graph's connectivity, as follows:

$$\mathbf{A}_{ij} = \begin{cases} w_{ij} & \text{if } (i,j) \in \mathcal{E} \text{ and } w_{ij} \geq \alpha \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The graph formulation process is visualized in Fig. 2, and the resulting graph includes a total of 53 nodes and 3904 pairs of edges, with $\alpha = 0.1$.

3.2 Feature Selection

From all the data collected in Section 2, we meticulously identify and integrate features integral to the graph formulation and the downscaling problem at hand, and form the augmented feature vector $\mathcal{X}$, serving as the input to our model. Specifically, for the i^{th} NoI, the feature vector $\mathcal{X}_i$ is composed of four groups of components,

$$\mathcal{X}_i = \left[\mathbf{x}_i^{WH}, \mathbf{x}_i^{AX}, \mathbf{x}_i^{GE}, \mathbf{x}_i^{GA}\right] \tag{3}$$

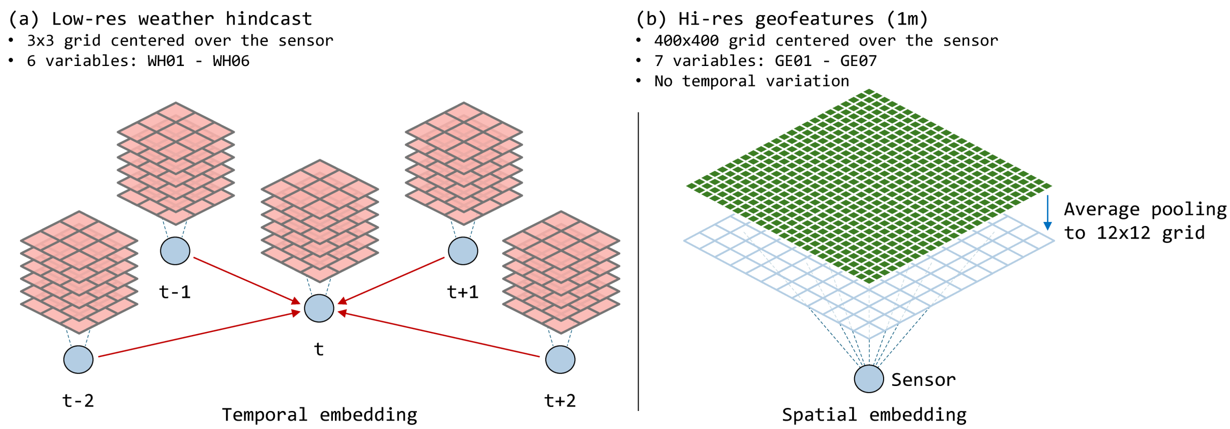

Figure 3: Illustration of feature assembly process for (a) variables from weather hindcast at 1km resolution; and (b) high-resolution geofeatures at 1m resolution.

Here, $\mathbf{x}_i^{WH}$ is the weather hindcast data (Section 2.1). Spatially, this component integrates the six surface meteorological variables, each from a 3×3 1km-grid centered at the NoI, thereby providing a general weather pattern over the NoI as well as its immediate vicinity. Temporally, it incorporates a 5-hour window (i. e., current, ± 2 time steps) across the 3×3 grid to inform the model with the temporal evolution of meteorological conditions. Figure 3a visualizes the assembly of WH variables, which are concatenated into a vector of $\mathbf{x}_i^{WH} \in \mathbb{R}^{6\times3\times3\times5}$ (6 variables in 3×3 grid with 5 time steps). Subsequently, the auxiliary group $\mathbf{x}_i^{AX} \in \mathbb{R}^3$ contains temporal metadata that is essential for the model prediction. Next, the $\mathbf{x}_i^{GE}$ component contains spatial embeddings consolidated from the seven geospatial maps (Fig. 1), each with an original resolution of 400×400 pixels. To facilitate a balance between preserving spatial details and ensuring the feature vector's manageability for the model, we apply a spatial averaging technique known as the average pooling. This is achieved by partitioning each map into smaller, non-overlapping subregions and calculating the average value within each subregion to represent its features. Consequently, this reduction technique transforms the original high-resolution data into a condensed format of 12×12 pixels for each of the seven maps (as shown in Fig. 3b), resulting in a composite feature vector of dimensions $\mathbf{x}_i^{GE} \in \mathbb{R}^{12\times12\times7}$. This approach allows us to maintain essential spatial information while ensuring the feature length remains concise, facilitating efficient processing by the model. Finally, $\mathbf{x}_i^{GA} \in \mathbb{R}^{20}$ include the statistical moments of the urban features without spatial or temporal embedding. For each NoI, the augmented vector $\mathcal{X}_i$ incorporates a total of 1302 features, with majority of information provided from embedding groups ($\mathbf{x}_i^{WH}$ and $\mathbf{x}_i^{GE}$). This comprehensive assembly ensures a rich amount of information for the ML model that facilitates an in-depth exploration of the urban climate dynamics.

4 Graph Neural Network

With the defined feature vector $\mathcal{X}^{(k)}$ and the collected street-level air temperature ($\mathbf{T}_a$) on all sensor at time step k, the downscaling question can be characterized as the following under GNN architecture:

$$\mathbf{T}_a^{(k)} = \mathbf{F}(\mathcal{X}^{(k)}, \mathcal{G}; \mathbf{\Theta}), \tag{4}$$

where the GNN model $\mathbf{F}$, parametrized by $\mathbf{\Theta}$, maps the extended state vector $\mathcal{X}^{(k)}$ (Eqn.(3)) at the current time step k to the street temperature $\mathbf{T}_a^{(k)}$, given the graph structure $\mathcal{G}$ of the AoT network. The GNN model used in this study is built upon the message passing (MP) mechanism, and utilizes an encoder-processor-decoder architecture. The key components of the proposed architecture are detailed as follows.

4.1 Message passing with GraphSAGE

The message passing (MP) mechanism serves as a foundational element across numerous Graph Neural Network (GNN) architectures, characterized by its execution of several consecutive MP steps. The GraphSAGE operator, introduced by Hamilton et al. exemplifies a spatial-based GNN designed to aggregate information from neighboring nodes (Hamilton et al., 2018). This operator is notable for its inductive framework that utilizes node attribute information to generate representations for previously unseen data efficiently.

Specifically, consider the graph representation denoted in Sec. 3.1 where each node $v \in \mathcal{V}$ has a node feature vector $\mathbf{h}_v \in \mathbb{R}^D$ and a set of neighbor nodes $u \in \mathcal{N}(v)$. At the j^{th} MP step, the new feature of node v is computed using its previous feature and information from its neighbors as,

$$\mathbf{m}^j_{\mathcal{N}(v)} = \text{AGGREGATE}\left(\{\mathbf{h}^j_u \mid u \in \mathcal{N}(v)\}, \mathcal{W}\right), \tag{5a}$$

$$\mathbf{h}^{j+1}_v = \text{UPDATE}\left(\mathbf{h}^j_v, \mathbf{m}^j_{\mathcal{N}(v)}, \mathcal{W}\right), \tag{5b}$$

where AGGREGATE denotes the aggregation scheme, e.g., mean aggregation, UPDATE are nonlinear mappings, e.g., neural networks, $\mathbf{m}_{\mathcal{N}(v)}$ denotes the information aggregated from the neighbors of node v, and $\mathcal{W}$ the set of trainable network parameters. One MP step corresponds to the information exchange between 1-hop neighbors, i.e., the nodes that directly connected. It is possible to stack multiple aggregators over k MP steps, and the feature vector of a node is influenced not only by its 1-hop neighbors, but also by the more distant k-hop neighbors.

4.2 Encoder-Processor-Decoder architecture

The GNN model uses an encoder-processor-decoder architecture that is shown in Fig. 4 and detailed as following.

1. Encoder: First, the encoder is applied to each individual node. It maps state vectors at a node $\mathbf{x}_i$, which consists of both continuous and discrete variables, to a latent vector $\mathbf{h}_i^0 \in \mathbb{R}^D$. The latent vector is a set of high-dimensional nonlinear features that provide a continuous representation of the states on each bus, which is amenable for NN computations. For the ith node at time step k, the encoder $\mathbf{f}_E$ is

$$\mathbf{h}_i^0 = \mathbf{f}_E(\mathbf{x}_i^{(k)}, \mathbf{x}_i^{(k-1)}, \cdots, \mathbf{x}_i^{(k-M+1)}; \mathbf{\Theta}^0), \tag{6}$$

where $\mathbf{f}_E$ is implemented as a standard fully-connected NN (FCNN) of N_M layers with a set of trainable parameters $\mathbf{\Theta}^0$. After the encoding, the latent vectors of all the nodes are denoted $\mathbf{H}^0 = \{\mathbf{h}_i^0\}_{i=1}^N \in \mathbb{R}^{N \times D}$.

2. Processor: Subsequently, a stack of $N = N_C$ graph MP layers serve as processors that successively aggregate the latent features from each node and its neighbors and update the latent vectors at each node. Formally, the j^{th} processor step is written as

$$\mathbf{H}^{j+1} = \mathbf{f}_P^j(\mathbf{H}^j; \mathbf{\Theta}^j), \tag{7}$$

where $\mathbf{f}_P^j$ is a GraphSAGE layer, with parameter $\mathbf{\Theta}^j$. In this case, N_C GraphSAGE layers are deployed to generate a series of the latent vectors $\mathbf{H}^1, \cdots, \mathbf{H}^{N_C}$ using (5a). The last output $\mathbf{H}^N$ is sent to the subsequent decoding step.

3. Decoder: Finally, the decoder maps the latent vector of each node to the desired output, i.e. the street level temperature,

$$\tilde{\mathbf{T}}_a^{(k)} = \mathbf{f}_D(\mathbf{H}^N; \mathbf{\Theta}^{N+1}), \tag{8}$$

where $\mathbf{f}_D$ is a FCNN of N_M layers with trainable parameters $\mathbf{\Theta}^{N+1}$.

4.3 Model implementation

Following the data preparation presented in Section 2 and Section 3.2, we compile a dataset comprising 2,944 hourly-recorded snapshots. From this dataset, a random selection of 70% is allocated for model training purposes, while the remaining is designated for validation and testing phases.

The GNN model is implemented using PyTorch Geometric (PyG) (Fey et al., 2019), an open-source

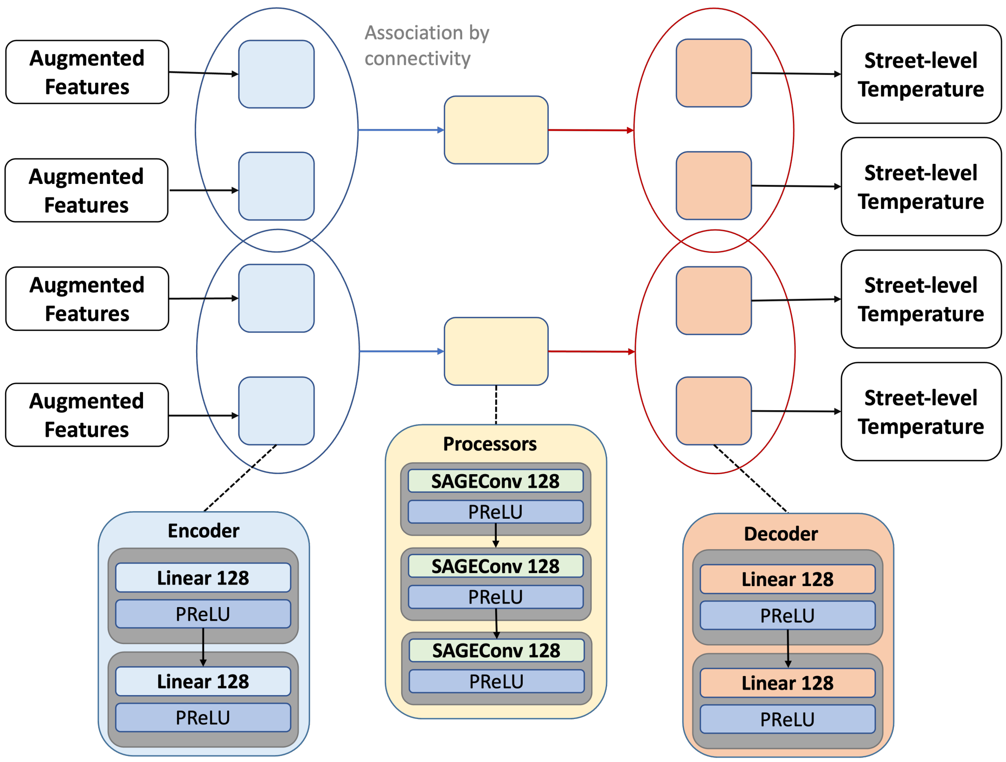

Figure 4: A structural diagram of the GNN model with encoder-processor-decoder architecture used in this study. This figure is redrew from Figure S3 in P. Li, Y. Yu, et al. (2023)).

machine learning framework with Graph Network architectures built upon PyTorch (Paszke et al., 2019). The size of latent vector (hidden dimension of the network) is chosen to be 128. The encoder and decoder modules each has two FCNN layers, and the processor is implemented with three GraphSAGE layers. Each layer of FCNN and GraphSAGE is followed with Parametric Rectified linear unit (PReLU) as the activation function (He et al., 2015). During training, the features and outputs of the model are normalized to a range of $[0, 1]$. To ensure the robustness of training, we use the Huber loss function with $\delta = 1.0$, which is minimized during training using the standard Adam optimizer (Kingma et al., 2017) with an exponential decay of learning rate.

4.4 Model Interpretation

Model interpretability refers to the ability to understand and articulate the internal mechanisms and decisions of a machine learning model (Murdoch et al., 2019). This understanding is crucial, as it enhances trust in the model's outputs by making the algorithm's processes transparent to end-users, especially in scenarios lacking ground truth. Furthermore, it illuminates the significance of various model features, e.g. how each patch of geofeatures is affecting the street-level temperatures. Understanding which features—such as green spaces or urban infrastructure—influence predictions the most can guide effective urban planning and climate mitigation strategies. However, most deep learning methods, traditionally designed for high performance rather than transparency, often lack inherent interpretability. We must then rely on post-hoc algorithms, which retrospectively analyze a trained ML model to identify and elucidate the factors influencing its decisions. These tools have become instrumental in uncovering the system's underlying knowledge, and in identifying critical features that significantly influence model outcomes, thereby offering valuable insights for informed decision-making and targeted urban planning initiatives.

GNNExplainer, introduced by Ying et al. (2019), is a post-hoc explanation algorithm tailored for GNN models. It aims to identify a compact, influential subgraph $\mathcal{G}_s$ and corresponding node features $\mathcal{X}_s$ that maximally preserve the prediction of the model. This is achieved through the maximization of mutual information (MI) between the predictions made using the original graph and those using the identified subgraphs. The mutual information is defined as:

$$\max_{\mathcal{G}_s} MI(\mathbf{Y}, (\mathcal{G}_s, \mathcal{X}_s)) = H(Y) - H(\mathbf{Y}|\mathcal{G} = \mathcal{G}_s, \mathcal{X} = \mathcal{X}_s) \tag{9a}$$

$$H(\mathbf{Y}) = -\int p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y} \tag{9b}$$

For a trained GNN model, the entropy $H(\mathbf{Y})$, where $p(\mathbf{y})$ is the probability of the model producing output $\mathbf{y}$, is constant when the model makes prediction with the complete graph. The maximization of MI is therefore the minimization of the conditional entropy $H(\mathbf{Y}|\mathcal{G} = \mathcal{G}_s, \mathcal{X} = \mathcal{X}_s)$, which computes for the expectation over the distribution of $\mathbf{Y}$ conditioned on the subgraph $\mathcal{G}_s$ and the corresponding node features $\mathcal{X}_s$.

By maximizing the mutual information between the predictions made using the original graph and the subgraph, it ensures that the subgraph captures the most important aspects of the original graph for the model's decision-making process. To identify $\mathcal{G}_s$, GNNExplainer applies a trainable soft mask $\mathbf{M}$ over the adjacency matrix $\mathbf{A}$, effectively adjusting edge weights to spotlight those pivotal for the model's decisions, thereby crafting a subgraph that maintains the predictive essence of the original graph. Besides providing explanations based on graph structures, GNNExplainer also extends its capabilities to feature-level insights by leveraging a similar soft-mask mechanism on node features, thereby generating normalized influence scores for each feature and offering a comprehensive understanding of both structural and feature-based contributions to the model's predictions.

In practice, GNNExplainer generates explanations by initially considering the entire graph and all features, then iteratively pruning edges and features that have the least effect on the prediction accuracy. This pruning is guided by gradient-based optimization techniques, which adjust the weights of edges and features to highlight those that contribute most significantly to the model's output. For an expansive explanation of the algorithm's workings and its application, we direct readers to the original work of Ying et al. (2019). Importantly, by using the true street-level temperature in computing Eqn.(9b), the GNNExplainer essentially elucidates the actual phenomena the model aims to capture, thereby offering a quantitative insight of how true street-level temperature is influenced by the various geofeatures, which are elaborated in Section 5.4.

4.5 Evaluation Metrics

The model performance is evaluated in three ways: (1) Overall performance, to provide a general accuracy and bias evaluation among all sensors as a system; (2) spatiotemporal distribution of model errors, to demonstrate the performance variances among different locations; (3) performance at out-of-sample locations, to test if the model can be generalized and quantify the uncertainties in prediction. Note that the model performance is only evaluated over the data points reserved for model validation. These validation data points are not used in model training.

The model performance is quantified using three metrics:

1. Root mean squared error (RMSE), defined as

$$RMSE = \sqrt{\frac{1}{tN} \sum_{k=1}^{t} \sum_{n=1}^{N} \left(\tilde{T}_{a,n}^{(k)} - T_{a,n}^{(k)} \right)^2}, \tag{10}$$

where the error between the predicted street temperature $\tilde{T}_a^{(k)}$ and true street temperature $T_a^{(k)}$ is averaged over total of t predictive time steps and N sensors.

2. Mean absolute error (MAE): This metric calculates the average magnitude of the errors in a set of predictions, without considering their direction. Compared to RMSE, which gives higher weight to large errors, MAE provides a more uniform measure of error magnitude.

$$MAE = \frac{1}{tN}\sum_{k=1}^{t}\sum_{n=1}^{N}\left|\tilde{T}_{a,n}^{(k)} - T_{a,n}^{(k)}\right|, \tag{11}$$

3. Mean bias error (MBE): This metric quantifies the average bias in the predictions, providing insight into whether the model tends to overestimate or underestimate the true values. It is calculated as:

$$MBE = \frac{1}{tN}\sum_{k=1}^{t}\sum_{n=1}^{N}(\tilde{T}_{a,n}^{(k)} - T_{a,n}^{(k)}), \tag{12}$$

A positive MBE indicates a tendency of the model to overestimate, while a negative value suggests an underestimation.

5 Results and Discussion

5.1 Model performance

For process-based models, RMSE between 2.0 °C and 2.5 °C is commonly acceptable over month-long simulations on air temperature at hourly intervals. Data-driven models generally have better performance, with RMSE ranging from 1 to 1.5 °C in existing studies (H. Wang et al., 2023). To better benchmark our GNN model to its implementation in Chicago, we replicate the Gaussian Process Regression method described in P. Li and A. Sharma (2024b) and train on the same labeled dataset used in this study as a reference.

As shown in Fig. 5a, the average RMSE of GNN model is 0.93 °C across the 53 sensors in the city of Chicago, which sits at the lower end of the spectrum of RMSE (i.e., 1 - 1.5 °C) for data-driven studies. The prediction accuracy is also better than the GPR model (1.21 °C, Fig. 3b). More importantly, GNN shows better consistency when predicting at different locations with a smaller standard deviation on sensor-wise RMSEs (0.06 °C GNN in Fig. 3c vs 0.25 °C GPR in Fig. 3d). Despite the overall improvements from GNN, it is intriguing that the RMSEs from these two algorithms exhibit a linear correlation with statistical significance (Fig. 3e). The convergence of their error patterns indicates their similar understanding and interpretation of the underlying data characteristics. The agreement in performance variances also implies the existence of favorable and unfavorable locations in general, which can guide further refinement of the models and the dataset. For instance, further feature engineering or data collection efforts should focus on those unfavorable locations. We will elaborate more on this point in Section 5.2.

Despite the performance of GNN is generally better, there are a few exceptions where GPR outperforms GNN (Fig. 5e). Certain sensors showing the highest RMSEs in GNN model are not necessarily the worst performer with GPR, *vice versa*. This variability is likely resulted from the inherent differences between the non-parametric nature of GPR and the parametric approach of GNN. The observed performance convergence and variability underscore the potential benefits of employing an ensemble of ML models by integrating multiple algorithms trained on the same dataset or slightly altered subset. Although the ensemble may not significantly enhance accuracy, it is anticipated to yield more reliable predictions and mitigate the risk of overfitting.

It is important to note that the deployment of the GNN in this research is not solely on outperforming the other modeling techniques, as each algorithm possesses its own unique advantages. Rather, our objective is to explore how GNN achieves superior results and to derive a more generalized, effective strategy for model selection, data organization, and the architectural design of ML models, particularly for simulating street-level dynamics. Beyond its enhanced accuracy, the GNN model demonstrates potential in unraveling the intricate interactions between geospatial locations, evidenced by its consistency across the space. Consequently, we anticipate the GNN model to provide more reliable predictions at out-of-sample locations, which is a critical factor in assessing model performance. We will further elaborate this in the next subsection.

5.2 Performance at out-of-sample locations

Theoretically, data extrapolation is a major challenge for all ML algorithms, meaning that ML models generally have worse performance on out-of-sample datasets. To further investigate the predictive capability of GNN model, we employ a "leave-one-sensor-out" (LOSO) testing strategy. This approach involves training a series of models, each excluding data from one specific sensor (P. Li and A. Sharma, 2024b). Compared to the model trained on labeled data from all sensors (hereinafter referred to as the nominal model) discussed in the previous section, each LOSO model is deprived of any information from the left-out sensor, thus can rigorously reflecting GNN's predictive accuracy on unfamiliar locations (i.e., out-of-sample locations). When evaluating their performances, RMSEs for nominal model and LOSO models will be calculated on the data points reserved for validation (i.e., out-of-sample data point). But since the nominal model has leveraged geofeatures from all sensors in training, it processes certain knowledge over all sensors. Conversely, when training LOSO models, the node and its associated edges corresponding to the left-out sensor are removed

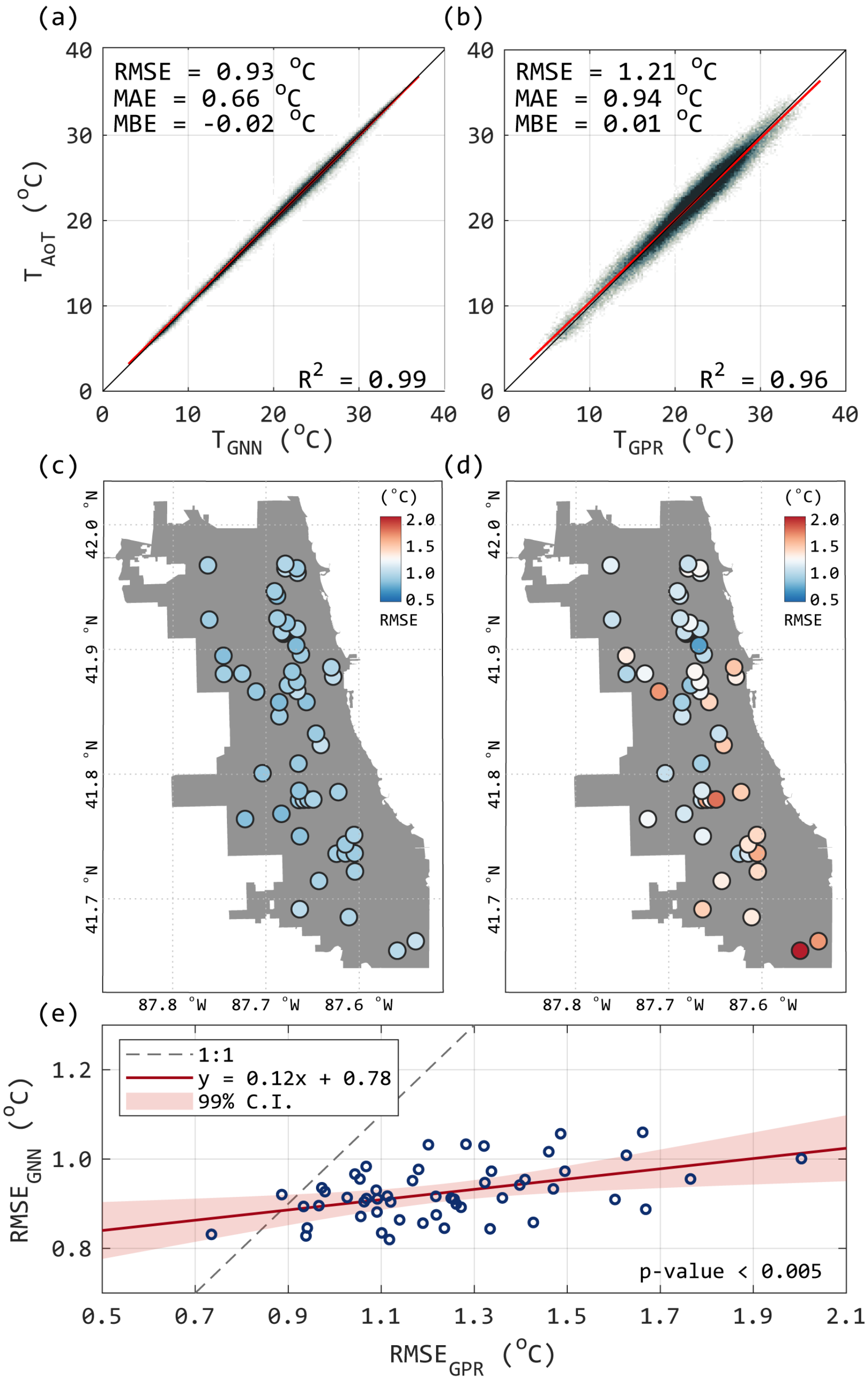

Figure 5: GNN model performance compared to GPR model presented in P. Li and A. Sharma (2024b). Scatter plot across simulation period (a) GNN; and (b) GPR. Spatial distribution of model RMSE (c) GNN; and (d) GPR. (e) Correlation of model performance between GNN and GPR at 53 sensor locations in Chicago. Red line in (e) is the linear regression between $\mathrm{RMSE_{GPR}}$ and $\mathrm{RMSE_{GNN}}$. Shaded zone in (e) indicates the 99% confidence interval of the linear regression.

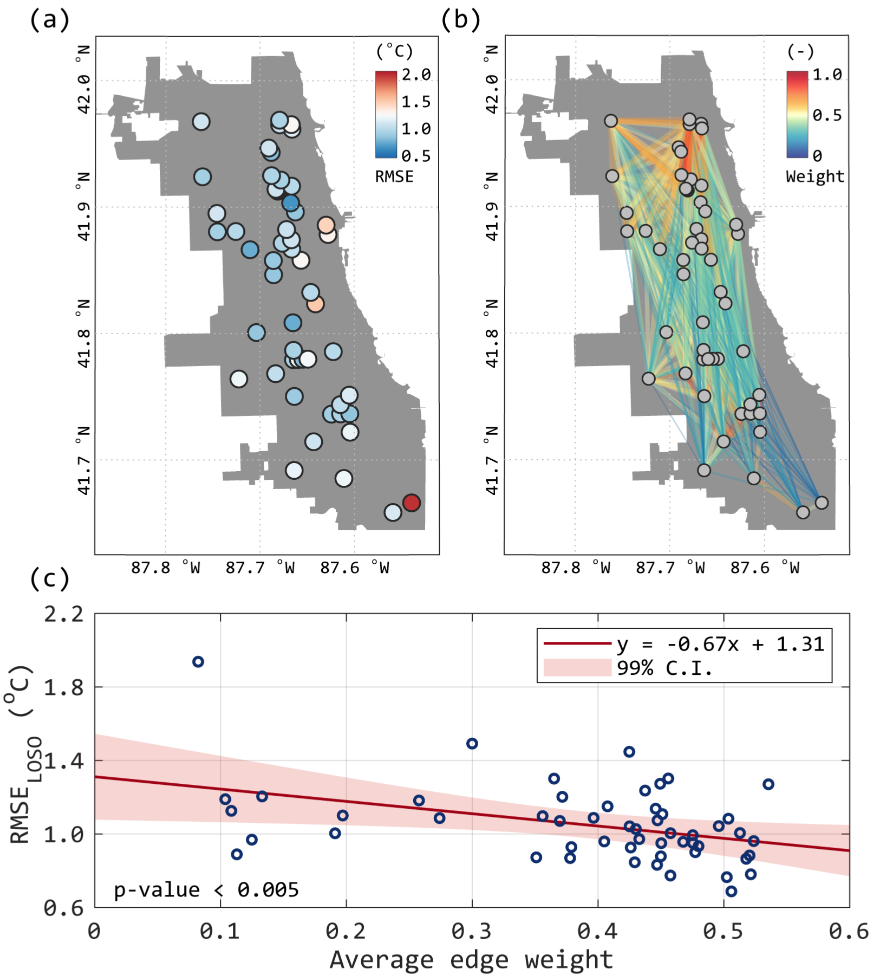

Figure 6: (a) Spatial distribution of the model performance with "leave-one-sensor-out" (LOSO) configuration. (b) Visualization of edge weight (w_{ij}) between paired nodes. (c) Correlation between averaged edge weight and RMSEs of LOSO models over 53 sensor locations.

from the graph. In the subsequent prediction phase, the corresponding node and edges will be incorporated as new information to the model. LOSO test mimics the practical processes of implementing a trained GNN model over any designated location in the city. It leverages GNN's intrinsic ability to adapt to graphs of varying topology, thereby ensuring the feasibility of predictions on new nodes.

Figure 6a shows the map of RMSE for LOSO test. As expected, the average performance of LOSO models is worse than the nominal model (1.03 °C vs 0.93 °C) with a greater variation over all sensors. But the average performance is still better than GPR model trained over all sensors, indicating GNN is more robust when making predictions on out-of-sample locations.

The discrepancies in performance might be originated from the difference in model architecture. GPR models predict the posterior distribution by incorporating prior knowledge and conditioning these predictions on provided geofeatures. Once certain geofeature is missing in the training dataset, GPR models must interpolate, or at times extrapolate, their impact on the target variable. This task can be challenging when the left-out geofeatures are unique across the locations. P. Li and A. Sharma (2024b) observed that the performance of GPR model can be improved significantly by including even a small subset of the measurement from the left-out sensor in training, underscoring the geofeatures' pivotal role in its modeling structure. In contrast, GNN models are not susceptible to this limitation. In addition to the geofeatures that are specific to each sensor, GNN model can learn the temporal evolution patterns from the training graph using its inductive framework. This capability is enhanced by the message passing mechanisms, which allow information exchange between the existing and new nodes depending on the assigned edge weights (Fig. 6b). Figure 6c shows the statistical relationship between LOSO model RMSEs (Fig. 6a) and averaged edge weight (Fig. 6b). We find that if the average edge weight of a sensor is higher (i.e., permitting more information exchange from the other sensors), its corresponding LOSO model will generally have a better performance. Since we use the distance and land use similarity between sensors to calculate edge weights (Section 3.1), a smaller edge weight value indicates the sensor is geographically isolated or unique in land use conditions, thus a likely worse predictive performance. This finding implies that the model performance can be improved by strategically selecting measurement locations that are close to each other or similar in land use. Practically, with limited number of sensors can be deployed, it would be helpful to distribute the sensors across the representative urban land covers with equal distancing. In fact, these information is encoded within the adjacency matrix as prior knowledge before training, therefore, the deployment locations can be derived by simply optimizing the adjacency matrix toward higher edge weight values across all planned sensors.

It is also noteworthy that the distinctions of GPR and GNN models make them specialized for different tasks. For example, GPR models will be more suitable for gap-filling on the timeseries at specific locations once the historical measurement is available. While GNN models will be more reliable for predictions over unseen locations. In this case, GNN is believed to be a promising approach to transfer the learned knowledge from one to the other cities. This capability can be extremely valuable as street-level observation networks are rather rare and can be time-consuming and labor-intensive to deploy, while datasets of geofeatures can be generated at a much more affordable cost.

5.3 Ablation test

In the previous section, we primarily benchmarked the GNN model and discussed its performance variances across different geospatial locations. The subsequent ablation study examines how the model's performance is affected by the absence of specific groups of input data, which will illustrates the impact of each variable group, and identify key contributors to the model performance within the established model architecture.

We first categorize the input features into four groups as shown in Eqn (3). The nominal model, discussed in previous sections, utilizes data from all categories (WH+GE+GA+AX). To assess the impact of each feature group on model performance, we prescribe three ablation models, (1) WH+GE+AX, excluding geofeatures calculated as statistical moments; (2) WH+GA+AX, omitting embedded geofeatures; and (3) WH+AX, where all geofeatures are removed. Table 2 summarizes the model configurations and performance. The findings indicate minimal performance variation when averaged geofeatures (GA) are excluded (comparing models 1 and 2, or models 3 and 4). Conversely, the inclusion of embedded geofeatures (GE) can significantly improve model performance, as evidenced by the comparisons between models 1 and 3, or models 2 and 4 in Table 2.

Table 2: Comparison of model performance in RMSE (°C)

Model No.	Configuration	Mean	Std	Best	Worst
1	Nominal model (WH + GE + GA + AX)	**0.92**	**0.06**	0.82	**1.06**
2	WH + GE + AX	1.06	0.08	0.90	1.26
3	WH + GA + AX	1.18	0.10	1.02	1.50
4	WH + AX	1.19	0.11	1.02	1.51
5	LOSO (WH + GE + GA + AX)	1.05	0.21	**0.69**	1.94
Ref	GPR	1.24	0.25	0.73	2.00

Despite the geofeatures in GE and GA group include similar data elements, such as land cover conditions and the heights of surface objects in vertical dimension, GE group offers an added dimension by detailing the spatial distribution of the geofeatures around the sensors. This granular information allows the model to quantify the significance of geofeatures based on their orientation relative to the sensors. For example, tall buildings in the upwind direction may largely alter the mixing condition of the street, thus having a stronger influence on street-level temperature (Gao et al., 2022). A similar situation applies to the localized shading effect from trees and buildings, which plays a major role in energy re-distribution in the built environment (Park et al., 2021; Wang, M and Yang, J., 2021).

The challenge in practice, though, is to assimilate the vast array of data (e.g., 1302 features use here) into a modeling framework, which proves to be daunting for certain ML algorithms such as the GPR. Due to its non-parametric nature, GPR model makes predictions based on every entry in the training dataset. Consequently, incorporating more data points or dimensions will lead to a cubic rise in computational complexity. This surge compromises efficiency in training and prediction, offsetting the advantages of using ML models for climate science. Conversely, the structure of GNN models can handle large datasets in a scalable and efficient manner, as its complexity depends on the predefined architecture, such as the number of hidden neurons and layers. This characteristic helps it remain manageable model size for applications with high-dimensional inputs.

5.4 Spatial pattern and feature significance

To assess the impact of geofeatures on street-level temperature more closely, we utilize GNNExplainer (Section 4.4) to compute the influence score for each variable within the GA and GE groups. Our analysis revealed significant variations in the importance of GA group geofeatures. Specifically, the fraction of impervious surfaces (GA16), vegetation (GA17), and Leaf Area Index (GA06) were identified the most influential geofeatures on street-level temperature. Conversely, water fractions (GA18) and high-intensity development (GA15) were found to be less impactful (Fig. 7a). When categorizing these geofeatures by their informational content, it becomes evident that planar land cover and land use attributes (e.g., fractions and development intensity) generally have higher influence scores than vertical measurements (e.g., building and canopy heights). This discrepancy likely stems from the relatively large averaging radius (200m) in comparison to the average height of surface objects in Chicago (<30m). This observation is consistent with findings from the GE group, where variables in horizontal dimension are deemed more critical than those in the vertical dimensions.

Another key discovery within the GA group is that canopy density variables hold a higher influence score than both canopy fraction (GA19) and canopy height (GA01-04) (Fig. 7a), despite being derived from datasets with coarser spatial resolution (300m for canopy density vs 1m for height information). Moreover, canopy density is the only geofeature group that has temporal variation over the summer months. Its higher influence score than the other static geofeatures indicates the critical impact of vegetation phenology on the hyper-local environment, even during a relatively short period. Yet, incorporating dynamic vegetation attributes in urban climate studies is uncommon, possibly due to the scarcity of accessible, city-specific vegetation data for modeling purposes. The absence of spatiotemporal vegetation data in high resolution also prevents us to include canopy density variables in GE group using the spatial embedding technique. This finding from our model and the present gap in data availability potentially imply a broader trend of underestimating the role of vegetation phenology in environmental modeling (Bernard et al., 2022; Zhou, 2022), even though research on the impact of urban heat on plant phenology is quite prevalent (Zipper et al., 2016; D. Li et al., 2019; Meng et al., 2020).

Analysis of the GE group supports observations from the GA group, revealing that geofeatures in horizontal dimension typically score higher than the vertical ones (Fig. 7b), meaning more influence on street-level temperature. In addition, we find that, compared to the NLCD classification, the CMAP data shows greater

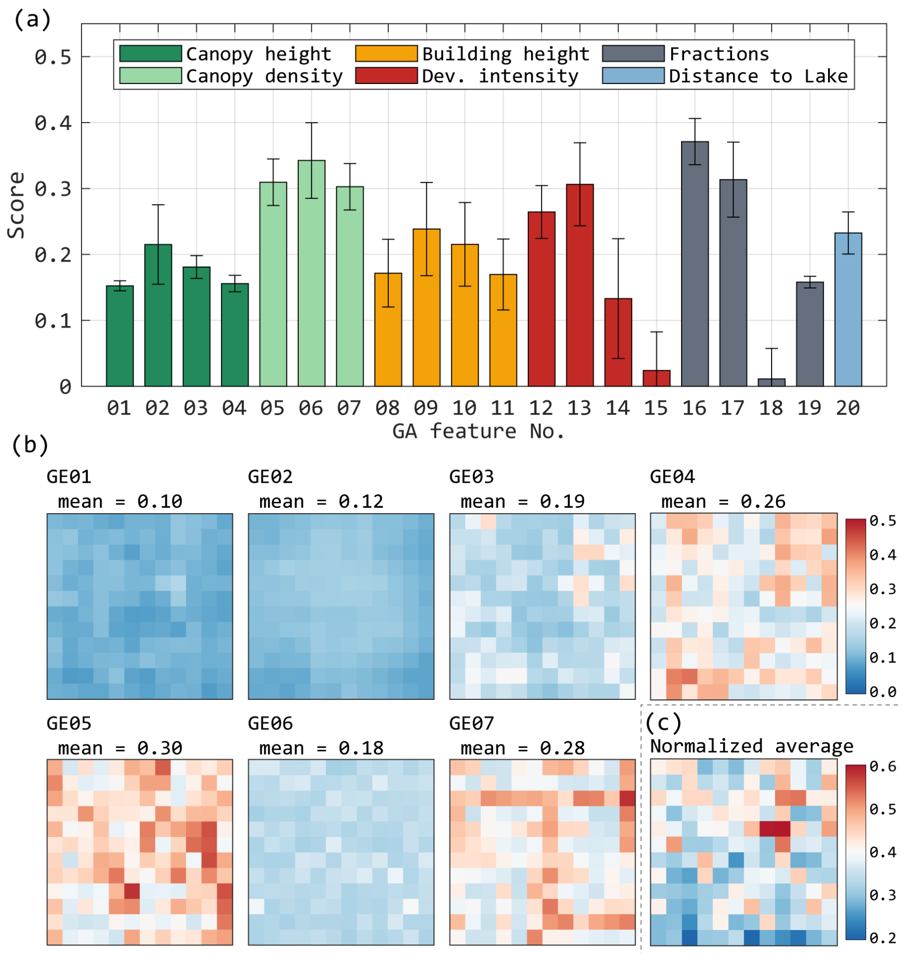

Figure 7: (a) Importance scores of individual features in GA group.(b) The spatial distribution of importance scores for embedded features in GE group. (c) The normalized average importance score of GE group.

significance. NLCD mainly categorizes urban land cover by the extent of impervious surfaces, offering limited insight into land use and building functionality. In contrast, Chicago's landscape, predominantly characterized by residential areas with medium-density housing and commercial centers, is oversimplified in NLCD's "medium development intensity" category. CMAP's data, with its higher spatial resolution and more nuanced urban classification, provides a more accurate depiction of land surface conditions. This enhanced characterization suggests model performance can benefit from a detailed description of urban land surfaces via a more representative classification scheme.

The spatial distribution of influence scores, derived from individual embedded features as shown in Fig. 7b, does not present a clear pattern. The GNN model makes predictions using specific location of the site, time of day, and day of the year, making it practically impossible to comprehend its mechanism at each timestep. However, aggregating features across all sensor locations reveals a discernible hotspot in the northeast direction (Fig. 7c), which intriguingly corresponds with the dominant wind direction (southwest to northeast) in Chicago during the summer. The spatial proximity of this hotspot to sensor locations, approximately 100 to 150m, coincides with distances identified in research seeking the optimal averaging radius for model efficiency and performance (Allen-Dumas et al., 2021). Given that high-resolution geospatial variables in the GE group are transformed into 12x12 matrices for the GNN model, pinpointing specific urban features responsible for this observation is challenging. Consequently, it is premature to draw definitive conclusions about spatial patterns of feature significance. Nonetheless, this suggests that employing more sophisticated embedding techniques (e.g. through an autoencoder) could illuminate the relationship between the layout of geofeatures and their thermal effects. Collectively, we advocate the development of comprehensive high-resolution urban climate informatics to help the investigation the microclimate dynamics via data-driven approach.

6 Concluding Remarks

In this study, we investigated the efficacy of Graph Neural Networks (GNN) in addressing the street-level downscaling problem at discrete locations, leading to four main contributions: (1) enhanced the precision of hourly air temperature predictions at the street level; (2) evaluated the model's ability on spatial extrapolation; and (3) examined how urban features influence street-level temperatures, thereby improving model interpretability and our understanding of microclimate dynamics; and (4) demonstrated the applicability of the hybrid modeling framework presented in (P. Li and A. Sharma, 2024b). Meanwhile, we compared the GNN model against the previous GPR model and digested their distinctions in architecture, data handling, and performance under various use cases. We concluded that the improve of prediction accuracy can be attributed to the architectural advantages of GNN and its capability of handing extensive high-dimensional datasets. Findings from model ablation and feature significance analysis elucidated the critical aspects of urban features, such as the dynamic canopy density data and detailed representative urban land classification, which can help to establish a nuanced benchmark for collecting environmental data in urban settings. It is also possible to use such modeling and analyzing methods to identify the dominance of physical processes at street-level microclimate based on the relative importance of all urban features. This can, in turn, inform the physics-based urban climate models to effectively focus on the predominant processes without introducing extra burdens on computation, thus promoting a synergistic cycle that enhances the Modeling – Experimenting (ModEx) strategy (DOE, 2020).

Along with its notable contributions, we reckon there are a few caveats of this study, which are not unique but rather common across contemporary data-driven urban climate research. These limitations highlight areas for future research efforts. One notable challenge is the lack of explicit dataset for anthropogenic heat sources in the model. The in-canyon thermal environment can be highly susceptible to anthropogenic heat sources from vehicles, buildings, and pedestrians. Though, to a certain degree, the spatial patterns of anthropogenic heat can be reflected from the land cover and land use and might be recognized by the ML model, the temporal variability is still underrepresented. As we concluded in Section 5.4, variables that change with time generally have higher importance than temporally static variables in the modeling process. This implies the criticalness to include the real-time traffic and building energy datasets to reflect the diurnal variations associated with rush hours and the difference between a weekday and weekend. Acquiring such hyper-local data across extensive areas presents significant challenges and, at times, may seem impractical without a direct application. Nevertheless, the validation of our framework and its novel application have implied the criticalness of these datasets, thereby justifying the effort to compile them for an in-depth investigation of anthropogenic heat's impact on hyper-local climates.

Besides the contributions and caveats from fundamental science perspective, insights from this study extend significantly into urban planning and policy. For example, by identifying key physical processes and urban features that influence microclimates, this study can inform targeted interventions to mitigate urban heat island effects, enhance urban resilience against climate change, and improve public health. The advocacy for enhanced urban data collection is contingent upon the establishment of comprehensive data policies and the support of robust cyberinfrastructures. Thus, we call upon the research community, urban planners, policymakers, and technology developers to engage in deeper collaboration. Collectively, we can push forward the agenda for sustainable urban development and environmental stewardship.

Acknowledgement

YY and DH are supported by National Science Foundation under the award DMS-2229435. PL and AS are supported the National Science Foundation awards 2139316, 2230772, and 2330565, and National Aeronautics and Space Administration award #80NSSC22K1683. This material is also based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research's Urban Integrated Field Laboratories CROCUS project research activity, under Contract Number DE-AC02-06CH11357. We would like to acknowledge high-performance computing supports from Cheyenne (doi: 10.5065/D6RX99HX) and Derecho (doi: 10.5065/qx9a-pg09) provided by NCAR's Computational and Information Systems Laboratory (CISL), sponsored by the National Science Foundation. We also acknowledge NOAA, the City of Chicago, and the Chicago Metropolitan Agency for Planning for providing the data used in this study.

Conflict of Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Research Statement

All the datasets used in this study are publicly available with open access and allow direct download. The Chicago land use inventory (LUI) can be found in CMAP (2023). AoT dataset can be found in ANL (2022). ILHMP LiDAR data can be found in ISGS (2019). GHCN dataset can be found in Menne et al. (2012). NLCD dataset can be found in Dewitz and USGS (2021).

References

Allen-Dumas, M., Sweet, L., and Brelsford, C. (2021). "Determining optimal resolution for urban terrain inputs to microclimate modeling". In: *ESS Open Archive*. DOI: 10.1002/essoar.10507082.1.

Bernard, É., Munck, C. de, and Lemonsu, A. (2022). "Detailed mapping and modeling of urban vegetation: What are the benefits for microclimatic simulations with Town Energy Balance (TEB) at neighborhood scale?" In: *Journal of Applied Meteorology and Climatology* 61, pp. 1159–1178. DOI: 10.1175/JAMC-D-21-0134.1.

Cao, Q. et al. (2022). "Modeling intra-urban differences in thermal environments and heat stress based on local climate zones in central Wuhan". In: *Building and Environment* 225, p. 109625. DOI: 10.1016/j.buildenv.2022.109625.

Catlett, C., Beckman, P., Sankaran, R., and Galvin, K. (2017). "Array of things: a scientific research instrument in the public way: platform design and early lessons learned". In: *Proceedings of the 2nd International Workshop on Science of Smart City Operations and Platforms Engineering*. SCOPE '17. Pittsburgh, Pennsylvania: Association for Computing Machinery, pp. 26–33.

Chen, B. et al. (2021). "Refined urban canopy parameters and their impacts on simulation of urbanization-induced climate change". In: *Urban Climate* 37, p. 100847. DOI: 10.1016/j.uclim.2021.100847.

Chen, F. and Dudhia, J. (2001). "Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity". In: *Monthly Weather Review* 129, pp. 569–585. DOI: 10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Chen, F., Kusaka, H., et al. (2011). "The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems". In: *International Journal of Climatology* 31, pp. 273–288. DOI: 10.1002/joc.2158.

Chen, Y., Wang, Y., Huang, G., and Tian, Q. (2024). "Coupling physical factors for precipitation forecast in China with Graph Neural Network". In: *Geophysical Research Letters* 51, e2023GL106676. DOI: 10.1029/2023GL106676.

Chen, Y.-C., Liao, Y.-J., Yao, C.-K., Honjo, T., Wang, C.-K., and Lin, T.-P. (2019). "The application of a high-density street-level air temperature observation network (HiSAN): The relationship between air temperature, urban development, and geographic features". In: *Science of The Total Environment* 685, pp. 710–722. DOI: 10.1016/j.scitotenv.2019.06.066.

Ching, J. et al. (2018). "WUDAPT: An urban weather, climate, and environmental modeling infrastructure for the Anthropocene". In: *Bulletin of the American Meteorological Society* 99, pp. 1907–1924. DOI: 10.1175/BAMS-D-16-0236.1.

Demuzere, M., Kittner, J., and Bechtel, B. (2021). "LCZ Generator: online tool to create Local Climate Zone maps". In: *EGU General Assembly 2021*, EGU21–11385. DOI: 10.5194/egusphere-egu21-11385.

DOE (2020). *The integrated model-experiment (ModEx) approach.* URL: https://ess.science.energy.gov/modex/#.

Ejurothu, P., Mandal, S., and Thakur, M. (2023). "Forecasting PM2.5 concentration in India using a cluster based hybrid Graph Neural Network approach". In: *Asia-Pacific Journal of Atmospheric Sciences* 59, pp. 545–561. DOI: 10.1007/s13143-022-00291-4.

Fey, M. and Lenssen, J. (2019). "Fast Graph Representation Learning with PyTorch Geometric". In: *ICLR Workshop on Representation Learning on Graphs and Manifolds*.

Gao, H., Liu, J., Lin, P., Li, C., Xiao, Y., and Hu, G. (2022). "Pedestrian level wind flow field of elevated tall buildings with dense tandem arrangement". In: *Building and Environment* 226, p. 109745. DOI: 10.1016/j.buildenv.2022.109745.

Gitahi, J. and Hahn, M. (2020). "High-resolution urban air quality monitoring using sentinel satellite images and low-cost ground-based sensor networks". In: *E3S Web of Conferences*. Vol. 171. E3S, p. 02002. DOI: 10.1051/e3sconf/202017102002.

Guo, R. et al. (2022). "High-resolution urban air quality mapping for multiple pollutants based on dense monitoring aata and machine learning". In: *International Journal of Environmental Research and Public Health* 19. DOI: 10.3390/ijerph19138005.

Hamilton, W., Ying, R., and Leskovec, J. (2018). *Inductive representation learning on large graphs.* arXiv: 1706.02216 [cs.SI].

Hammerberg, K., Brousse, O., Martilli, A., and Mahdavi, A. (2018). "Implications of employing detailed urban canopy parameters for mesoscale climate modelling: a comparison between WUDAPT and GIS databases over Vienna, Austria". In: *International Journal of Climatology* 38, e1241–e1257. DOI: 10.1002/joc.5447.

He, H., Zhang, Z., Ren, S., and Sun, J. (2015). *Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification.* arXiv: 1502.01852 [cs.CV].

Hong, S.-Y. and J.-O.J., L. (2006). "The WRF Single-Moment 6-Class Microphysics Scheme (WSM6)". In: *Asia-pacific Journal of Atmospheric Sciences* 42, pp. 129–151. URL: https://api.semanticscholar.org/CorpusID:120362377.

Hong, S.-Y., Noh, Y., and Dudhia, J. (2006). "A new vertical diffusion package with an explicit treatment of entrainment processes". In: *Monthly Weather Review* 134, pp. 2318–2341. DOI: 10.1175/MWR3199.1.

Iacono, M., Delamere, J., Mlawer, E., Shephard, M., Clough, S., and Collins, W. (2008). "Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models". In: *Journal of Geophysical Research: Atmospheres* 113, p. D13103. DOI: 10.1029/2008JD009944.

Jin, L., Schubert, S., Fenner, D., Meier, F., and Schneider, C. (2021). "Integration of a building energy model in an urban climate model and its application". In: *Boundary-Layer Meteorology* 178, pp. 249–281. DOI: 10.1007/s10546-020-00569-y.

Kain, J. (2004). "The Kain–Fritsch convective parameterization: An update". In: *Journal of Applied Meteorology* 43, pp. 170–181. DOI: 10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.

Keisler, R. (2022). *Forecasting global weather with Graph Neural Networks.* arXiv: 2202.07575.

Kim, M., Jeong, D., and Kim, Y. (2021). "Local climate zone classification using a multi-scale, multi-level attention network". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 181, pp. 345–366. DOI: 10.1016/j.isprsjprs.2021.09.015.

Kingma, D. and Ba, J. (2017). *Adam: A method for stochastic optimization.* arXiv: 1412.6980 [cs.LG].

Kondo, H., Genchi, Y., Kikegawa, Y., Ohashi, Y., Yoshikado, H., and Komiyama, H. (2005). "Development of a multi-layer urban canopy model for the analysis of energy consumption in a big city: Structure of the urban canopy model and its basic performance". In: *Boundary-Layer Meteorology* 116, pp. 395–421. DOI: 10.1007/s10546-005-0905-5.

Kousis, I., Pigliautile, I., and Pisello, A. (2021). "Intra-urban microclimate investigation in urban heat island through a novel mobile monitoring system". In: *Scientific Reports* 11, p. 9732. DOI: 10.1038/s41598-021-88344-y.

Krayenhoff, E. et al. (2020). "A multi-layer urban canopy meteorological model with trees (BEP-Tree): Street tree impacts on pedestrian-level climate". In: *Urban Climate* 32, p. 100590. DOI: 10.1016/j.uclim.2020.100590.

Lam, R. et al. (2023). "Learning skillful medium-range global weather forecasting". In: *Science* 382, pp. 1416–1421. DOI: 10.1126/science.adi2336.

Li, D., Stucky, B., Deck, J., Baiser, B., and Guralnick, R. (2019). "The effect of urbanization on plant phenology depends on regional temperature". In: *Nature Ecology & Evolution* 3, pp. 1661–1667. DOI: 10.1038/s41559-019-1004-1.

Li, P. and Sharma, A. (2024a). "HiTAB-Chicago: HeIght map of Trees And Buildings for the city of Chicago". In: *Zenodo.* DOI: 10.5281/zenodo.10463648.

— (2024b). "Hyper-local temperature prediction using detailed urban climate informatics". In: *Journal of Advances in Modeling Earth Systems* 15, e2023MS003867. DOI: 10.1029/2023MS003867.

Li, P., Sharma, A., Wang, Z.-H., and Wuebbles, D. (2023). "Assessing impacts of environmental perturbations on urban biogenic carbon exchange in the Chicago region". In: *Journal of Advances in Modeling Earth Systems* 15.10, e2023MS003867. DOI: 10.1029/2023MS003867.

Li, P. and Wang, Z.-H. (2020). "Modeling carbon dioxide exchange in a single-layer urban canopy model". In: *Building and Environment* 184, p. 107243. DOI: 10.1016/j.buildenv.2020.107243.

Li, P., Yu, Y., Huang, D., Wang, Z.-H., and Sharma, A. (2023). "Regional heatwave prediction using Graph Neural Network and weather station data". In: *Geophysical Research Letters* 50, e2023GL103405. DOI: 10.1029/2023GL103405.

Lira, H., Martí, L., and Sanchez-Pi, N. (2022). "A Graph Neural Network with spatio-temporal attention for multi-sources time series data: An application to frost forecast". In: *Sensors* 22. DOI: 10.3390/s22041486.

Ma, M. et al. (2023). "HiSTGNN: Hierarchical spatio-temporal graph neural network for weather forecasting". In: *Information Sciences* 648, p. 119580. DOI: 10.1016/j.ins.2023.119580.

Meili, N. et al. (2020). "An urban ecohydrological model to quantify the effect of vegetation on urban climate and hydrology (UT&C v1.0)". In: *Geoscientific Model Development* 13, pp. 335–362. DOI: 10.5194/gmd-13-335-2020.

Meng, L. et al. (2020). "Urban warming advances spring phenology but reduces the response of phenology to temperature in the conterminous United States". In: *Proceedings of the National Academy of Sciences* 117, pp. 4228–4233. DOI: 10.1073/pnas.1911117117.

Meyer, D. et al. (2020). "WRF-TEB: Implementation and evaluation of the coupled Weather Research and Forecasting (WRF) and Town Energy Balance (TEB) model". In: *Journal of Advances in Modeling Earth Systems* 12. e2019MS001961 10.1029/2019MS001961, e2019MS001961. DOI: 10.1029/2019MS001961.

Middel, A., Nazarian, N., Demuzere, M., and Bechtel, B. (2022). "Urban climate informatics: An emerging research field". In: *Frontiers in Environmental Science* 10. DOI: 10.3389/fenvs.2022.867434.

Murdoch, W., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. (2019). "Definitions, methods, and applications in interpretable machine learning". In: *Proceedings of the National Academy of Sciences* 116, pp. 22071–22080. DOI: 10.1073/pnas.1900654116.

Myhre, G. et al. (2019). "Frequency of extreme precipitation increases extensively with event rareness under global warming". In: *Scientific Reports* 9. DOI: 10.1038/s41598-019-52277-4.

Oke, T. (2008). *Urban observations. Guide to meteorological instruments and methods of observation, Part II of Observing Systems*. URL: https://www.weather.gov/media/epz/mesonet/CWOP-WMO8.pdf.

Oke, T., Mills, G., Christen, A., and Voogt, J. (2017). *Urban Climates*. DOI: 10.1017/9781139016476.

Park, Y., Guldmann, J.-M., and Liu, D. (2021). “Impacts of tree and building shades on the urban heat island: Combining remote sensing, 3D digital city and spatial regression approaches”. In: *Computers, Environment and Urban Systems* 88, p. 101655. DOI: 10.1016/j.compenvurbsys.2021.101655.

Paszke, A. et al. (2019). “PyTorch: An imperative style, high-performance deep learning library”. In: *Advances in Neural Information Processing Systems 32*. Ed. by Wallach, H., Larochelle, H., Beygelzimer, A., d’Alché-Buc, F., Fox, E., and Garnett, R. Curran Associates, Inc., pp. 8024–8035.

Perkins-Kirkpatrick, S. and Lewis, S. (2020). “Increasing trends in regional heatwaves”. In: *Nature Communications* 11, p. 3357. DOI: 10.1038/s41467-020-16970-7.

Pilant, A., Endres, K., Rosenbaum, D., and Gundersen, G. (2020). “US EPA EnviroAtlas Meter-Scale Urban Land Cover (MULC): 1-m pixel land cover class definitions and guidance”. In: *Remote Sensing* 12. DOI: 10.3390/rs12121909.

Ribeiro, I., Martilli, A., Falls, M., Zonato, A., and Villalba, G. (2021). “Highly resolved WRF-BEP/BEM simulations over Barcelona urban area with LCZ”. In: *Atmospheric Research* 248, p. 105220. DOI: 10.1016/j.atmosres.2020.105220.

Scarselli, F., Gori, M., Tsoi A.C.and Hagenbuchner, M., and Monfardini, G. (2009). “The Graph Neural Network model”. In: *IEEE Transactions on Neural Networks* 20, pp. 61–80. DOI: 10.1109/TNN.2008.2005605.

Scott, A., Zaitchik, B., Waugh, D., and O’Meara, K. (2017). “Intraurban temperature variability in Baltimore”. In: *Journal of Applied Meteorology and Climatology* 56, pp. 159–171. DOI: 10.1175/JAMC-D-16-0232.1.

Sharma, A., Fernando, H., Hamlet, A., Hellmann, J., Barlage, M., and Chen, F. (2017). “Urban meteorological modeling using WRF: a sensitivity study”. In: *International Journal of Climatology* 37, pp. 1885–1900. DOI: 10.1002/joc.4819.

Sharma, S., Singh, A., Kumar, R., and Mehta, A. (2023). “Analysis of land surface temperature at 10-meter pixel size (spatial resolution) for Ahmedabad City”. In: *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, pp. 134–138. DOI: 10.1109/CISES58720.2023.10183518.

Silverman, A. et al. (2022). “Making waves: Uses of real-time, hyperlocal flood sensor data for emergency management, resiliency planning, and flood impact mitigation”. In: *Water Research* 220, p. 118648. DOI: 10.1016/j.watres.2022.118648.

Singh, M. et al. (2023). “DownScaleBench for developing and applying a deep learning based urban climate downscaling- first results for high-resolution urban precipitation climatology over Austin, Texas”. In: *Computational Urban Science* 3, p. 22. DOI: 10.1007/s43762-023-00096-9.

Skamarock, W. et al. (2021). "A description of the Advanced Research WRF Model version 4.3 (No. NCAR/TN-556+STR)". In: *NCAR Technical Notes*. DOI: doi:10.5065/1dfh-6p97.

Stavropulos-Laffaille, X. et al. (2018). "Improvements to the hydrological processes of the Town Energy Balance model (TEB-Veg, SURFEX v7.3) for urban modelling and impact assessment". In: *Geoscientific Model Development* 11, pp. 4175–4194. DOI: 10.5194/gmd-11-4175-2018.

Stewart, I. and Oke, T. (2012). "Local Climate Zones for urban temperature studies". In: *Bulletin of the American Meteorological Society* 93, pp. 1879–1900. DOI: 10.1175/BAMS-D-11-00019.1.

Tien, I., Lozano, J.-M., and Chavan, A. (2023). "Locating real-time water level sensors in coastal communities to assess flood risk by optimizing across multiple objectives". In: *Communications Earth & Environment* 4, p. 96. DOI: 10.1038/s43247-023-00761-1.

United Nations (2019). *World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)*. New York.

Venter, Z., Barton, D., Gundersen, V., Figari, H., and Nowell, M. (2020). "Urban nature in a time of crisis: recreational use of green space increases during the COVID-19 outbreak in Oslo, Norway". In: *Environmental Research Letters* 15, p. 104075. DOI: 10.1088/1748-9326/abb396.

Wang, A. et al. (2023). "Hyperlocal environmental data with a mobile platform in urban environments". In: *Scientific Data* 10, p. 524. DOI: 10.1038/s41597-023-02425-3.

Wang, C., Zhu, Y., Zang, T., Liu, H., and Yu, J. (2021). "Modeling inter-station relationships with attentive temporal Graph Convolutional Network for air quality prediction". In: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. WSDM '21. Virtual Event, Israel: Association for Computing Machinery, pp. 616–634. DOI: 10.1145/3437963.3441731.

Wang, F., Tian, D., Lowe, L., Kalin, L., and Lehrter, J. (2021). "Deep Learning for Daily Precipitation and Temperature Downscaling". In: *Water Resources Research* 57, e2020WR029308. DOI: 10.1029/2020WR029308.

Wang, H., Yang, J., Chen, G., Ren, C., and Zhang, J. (2023). "Machine learning applications on air temperature prediction in the urban canopy layer: A critical review of 2011–2022". In: *Urban Climate* 49, p. 101499. DOI: 10.1016/j.uclim.2023.101499.

Wang, J., Liu, Z., Foster, I., Chang, W., Kettimuthu, R., and Kotamarthi, V. (2021). "Fast and accurate learned multiresolution dynamical downscaling for precipitation". In: *Geoscientific Model Development* 14, pp. 6355–6372. DOI: 10.5194/gmd-14-6355-2021.

Wang, J., Qian, Y., et al. (2023). "Contrasting effects of lake breeze and urbanization on heat stress in Chicago metropolitan area". In: *Urban Climate* 48, p. 101429. DOI: 10.1016/j.uclim.2023.101429.

Wang, S., Li, Y., Zhang, J., Meng, Q., Meng, L., and Gao, F. (2020). "PM2.5-GNN: A domain knowledge enhanced Graph Neural Network for PM2.5 forecasting". In: *Proceedings of the 28th International Confer-*

ence on Advances in Geographic Information Systems. SIGSPATIAL '20. Seattle, WA, USA: Association for Computing Machinery, pp. 163–166. DOI: 10.1145/3397536.3422208.

Wang, M and Yang, J. (2021). "The impact of building height on urban thermal environment in summer: A case study of Chinese megacities". In: *PLoS One* 16. DOI: 10.1371/journal.pone.0247786.

Ying, Z., Bourgeois, D., You, J., Zitnik, M., and Leskovec, J. (2019). "GNNExplainer: Generating explanations for Graph Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. Vol. 32. Curran Associates, Inc.

Yu, D. and Fang, C. (2023). "Urban remote sensing with spatial big data: A review and renewed perspective of urban studies in recent decades". In: *Remote Sensing* 15. DOI: 10.3390/rs15051307.

Zhou, Y. (2022). "Understanding urban plant phenology for sustainable cities and planet". In: *Nature Climate Change* 12, pp. 302–303. DOI: 10.1038/s41558-022-01331-7.

Zipper, S., Schatz, J., Singh, A., Kucharik, C., Townsend, P., and Loheide, S. (2016). "Urban heat island impacts on plant phenology: intra-urban variability and response to land cover". In: *Environmental Research Letters* 11, p. 054023. DOI: 10.1088/1748-9326/11/5/054023.