ARC Alliance

ARC Geophysical Research

GR Geophysical Research

Hourly Urban Air Temperature Forecasting with Graph Machine Learning

Christophe Roger[1], Martin Hendrick[1], Moritz Burger[2,3], Nadav Peleg[4,5], Simone Fatichi[6], and Gabriele Manoli[*1]

[1]Laboratory of Urban and Environmental Systems, École Polytechnique Fédérale de Lausanne

[2]Institute of Geography, University of Bern

[3]Oeschger Centre for Climate Change Research, University of Bern

[4]Institute of Earth Surface Dynamics, University of Lausanne

[5]Expertise Center for Climate Extremes, University of Lausanne

[6]Department of Civil and Environmental Engineering, National University of Singapore

Abstract

The Urban Heat Island (UHI) effect, where urban areas exhibit higher temperatures than their rural surroundings, is a growing subject of concern due to its implications for human health, energy demand, and anthropogenic emissions. Accurate high-resolution forecasts of the UHI intensity and, more generally, of urban air temperatures are therefore crucial for guiding mitigation and adaptation strategies, especially for real-time heat warning systems and reliable power load forecasting. We present a spatiotemporal graph-based machine-learning framework for hourly urban air-temperature forecasting that couples a Diffusion Convolutional Recurrent Neural Network (DCRNN) encoder with a multi-horizon Multi-Layer Perceptron (MLP) decoder. The model is trained and evaluated using a dense network of 113 low-cost temperature sensors in Bern, Switzerland. Spatial dependencies are learned on a directed, weighted sensor graph built from geographic proximity and environmental similarity, while temporal dynamics are modeled with gated recurrence. We forecast temperatures at each sensor up to 24 hours ahead and compare two settings: conditioning the decoder on past rural-station meteorological observations and future meteorological forecasts. At 24-hour lead time, the proposed model achieves an average RMSE of 2.99 K with past meteorology and 1.68 K with future meteorology. Relative to a per-sensor RNN baseline, it improves performance by an average of 13% when only past meteorological data are available, but underperforms when future observations are provided, motivating further work on how to incorporate future exogenous variables. Finally, we demonstrate the practical value of the approach by combining the sensor forecasts with regression-kriging to produce hourly 50 m resolution temperature and UHI-intensity maps over Bern. Overall, the results show the promise of graph-based learning for city-scale, high-resolution temperature forecasting to support heat-risk management and urban planning.

*Corresponding Author

Keywords: Urban Heat Island, Time-Series Forecasting, Graph Machine Learning, Urban Air Temperature, Early Warnings

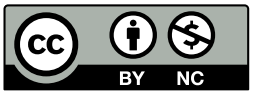
1 Introduction

Due to climate change, global mean temperatures are rising across the globe [40] and the frequency and intensity of heatwaves are increasing [76]. This raises concerns for the health and well-being of a growing population, especially in cities where heat stress and exposure are amplified [39]. Urban areas are generally more exposed to thermal stress due to the Urban Heat Island (UHI) effect [e.g., 55, 68], which results in higher urban temperatures than in the surrounding rural areas. The UHI effect is due to urban-induced modifications of land-surface properties and the surface energy balance [e.g., 44, 59, 71] and can lead to higher health risks, such as increased heat-related mortality [39, 82], higher energy demand due to air conditioning consumption and inaccurate electricity demand forecasting [22, 46, 49], and may even extend to impacting precipitation patterns around cities [16, 85].

Since the seminal work of Howard in the early 19th century [37], urban climate research has significantly advanced our understanding of the drivers, physical principles, and dynamic behavior of UHIs around the world. From a modeling perspective, several approaches have been developed to describe the role of built surfaces in weather forecasting models, from simple slab schemes to multi-layer canopy and building-resolving models [53]. However, urban land-surface models generally simulate temperature fields at a relatively coarse spatio-temporal resolution (e.g., 1-2 km and 1 h) [53], thus limiting their ability to simulate urban microclimate (e.g., at the street-scale). This can be achieved through computational fluid dynamics (CFD) approaches, resolving heat, mass, and momentum conservation equations at sub-meter scales [e.g. 29, 48, 56, 58, 61]. Yet, even if CFD models can now be applied to large urban domains (e.g., city scale), their usability in operational contexts remains limited due to their computational burden. Therefore, to address these challenges, computationally efficient alternatives are essential, with data-driven models emerging as particularly promising. These methods leverage extensive climate data obtained from ground-based monitoring stations or remote sensing platforms, integrating them with advanced statistical and machine learning techniques to model and predict the spatio-temporal dynamics of urban air temperatures at high resolution.

In this context, the growing availability of urban climate data, in tandem with the rapid advances in machine learning techniques, provides a valuable alternative to traditional physically-based approaches to assess the space-time dynamics of urban microclimate at resolutions as high as 1 m [2, 5, 7, 14, 69]. Statistical approaches, such as land-use regression modeling [e.g., 10, 77] or multiple linear regression [e.g., 43, 80], have been largely used to simulate the spatial patterns of UHIs. In recent years, the use of advanced machine learning (ML) methods has further surged [28, 89] to capture complex non-linear relationships between a large set of predictors and the UHI spatio-temporal patterns [e.g., 17, 23, 31]. Advanced ML methods have the advantage of being able to model subtle non-linear relationships while being considerably lower in computational cost compared to numerical weather forecasting and CFD models, which require solving a large set of non-linear differential equations [6, 60, 78]. Already in 1999, Santamouris et al. [73] used a neural network approach to

E-mail address: gabriele.manoli@epfl.ch
doi:10.5149/ARC-GR.1967



model the UHI phenomenon using ambient air temperature records from 20 locations across the city of Athens. With the development of more sophisticated machine learning methods, like tree-based models, recurrent neural networks (RNN), and convolutional neural networks (CNN), urban air temperature models were further advanced. In a recent review on the topic, Wang et al. [89] observed that tree-based methods [36, 83, 94, 102] perform well in spatial temperature prediction, although they fail to predict extreme values and are not suitable for temporal-based tasks because each sample is treated independently of the others. On the other hand, neural network-based methods (e.g., RNN, CNN) can be much more flexible and adaptable to various configurations and can handle non-linearity better than tree-based models. However, they are computationally more expensive to train, less interpretable (black-box models), and prone to overfitting [17, 35, 83, 94]. Despite these advances, Wang et al. [89] highlight a persistent gap in the field: most studies predict air temperature only at a few selected lead times (e.g., 1-h, 4-h, or 24-h), so it remains unclear how well models can handle continuous hour-by-hour multistep forecasts.

Many UHI-ML studies rely on remote sensing observations but such data are often limited to surface temperature observations during specific times of day, they are affected by cloud cover, and often contain errors of 1–2K [25, 51, 89]. The increasing deployment of ground-based sensors, such as crowdsourced networks like the Personal Weather Station Network [84], offers the potential to improve the spatio-temporal description of urban climate [19], especially in the context of data-driven modeling. Such high-resolution in-situ datasets are particularly attractive for the development of graph neural networks, a novel method for processing temporal observations from a sparse network of monitoring sensors. Graph neural networks (GNN) [74] enable the processing of any type of data that can be represented in a network format (also known as non-Euclidean data) [74, 92]. They have been used for numerous spatio-temporal prediction tasks, including traffic forecasting [42, 81, 96, 101], power load forecasting [41, 52, 64], and air quality prediction [26, 57, 70]. Yet, the use of GNNs for urban air temperature prediction is relatively underexplored. Yu et al. [98] developed a GNN with GraphSAGE layers [33] using a network of sensors with embedded spatial features and past and future weather features in the city of Chicago. Yu et al. [97] used graph attention layers [86, 87] coupled with gated recurrent units to predict temperature at several locations in China. In both cases, the GNN methods provided comparable or better results than other baseline approaches [89], such as Gaussian Process Regression [47] and Long-Short Term Memory [90].

Here, we develop a GNN method to further explore and improve continuous hour-by-hour predictions of urban air temperatures at high space and time resolutions. Using the city of Bern (Switzerland) as a case study, we make hour-by-hour, multi-step predictions of local UHI intensities with a 24-hour time horizon, mapping it for different forecasting time horizons at 50 m resolution. We test two scenarios and compare the results with a traditional RNN model: (i) a baseline scenario where GNN predictions are made using past weather information only (i.e., we assume no regional forecasts are available); and (ii) a forecasting scenario where future weather information at a reference station are available. Our GNN approach allows us to make high-resolution, one day-ahead forecasts of UHI intensity at the city scale, which could be used to issue early warnings in case of extreme heat or to obtain more accurate power load forecasts in urban areas.

2 Data and Methods

In the following, we present the study area, the data collection process, its organization into a graph-ready dataset, and the architecture and training methods of the proposed graph machine learning model.

2.1 Study Area

The study focuses on the city of Bern, Switzerland, and its surrounding metropolitan area with a total population size of approximately 240,000 people [9]. In 2018, the University of Bern installed a low-cost measurement network to study intra-urban temperature differences [10]. The network is composed of 113 self-built low-cost devices that measure air temperature at 10-minute intervals [32]. Records were collected and compiled for 6 consecutive years (2019-2024); during the first five years, the measurements were only available during the warm season (May 15th to September 15th), while from May 2023 on, continuous data is available until September 2024. Data from the sensor network was used to map daily urban heat island intensities [12] and model the spatial patterns of heatwaves in earlier studies [10].

2.1.1 Spatial Features

The spatial feature set encapsulates the static characteristics of the urban environment that have been chosen to describe how heat is stored, emitted, and redistributed across urban surfaces. Urban morphology metrics, such as average building height, building count, and construction density within a 100 m radius from each sensor, capture the complexity of the urban fabric, influencing airflow patterns, shading, and heat exchanges [66]. Population density [59] and land cover classifications based on Local Climate Zones [79] reflect anthropogenic activities and surface materials, while vegetated areas usually moderate urban-induced warming through evapotranspiration [30, 100]. Elevation and proximity measures (distance to the city center, corresponding to the most densely populated area of Bern, or to green spaces) further amplify temperature gradients, as higher-altitude neighborhoods generally experience cooler temperatures and parks serve as localized cool islands [13], although in Bern, due to the complex topography of the city, this might not always be the case. These spatial predictors encode the environmental and spatial context of each sensor, forming the basis for the sensor graph construction and the subsequent urban air temperature forecasting. Figure 1 illustrates the study area characteristics. The list of spatial features is provided in Table A.1.

2.1.2 Temporal Features

Temporal features provide essential dynamic context that drives the diurnal and seasonal variability of urban temperatures and UHI intensity under different meteorological conditions. Generally, UHI intensity (when defined in terms of air temperature, as done here) is more pronounced during nighttime compared to daytime [88], and in the case of Bern, it is particularly high on clear sunny days relative to other weather conditions [10, 11]. These temporal features include hour and month, encoding diurnal and seasonal cycles, capturing sunrise and sunset heating/cooling transitions as well as summer–winter contrasts in solar angle and day length. Meteorological variables from MeteoSwiss [1], including humidity, wind speed, incoming solar radiation, and atmospheric pressure, define the instantaneous meteorology at a reference station. These time series (summarized in Table A.2) provide the past 24 hours context on the weather dynamics and are used only during the prediction step. This allows testing the ability of the GNN model to predict future conditions with knowledge of past weather only. To test a more realistic forecasting application, we run an additional simulation experiment where, instead of past observations, we include future observations from a reference station in the rural area (for the next 24 hours) in the prediction step. In this case, actual observations are used as a proxy for (coarser) regional weather forecasts in order to avoid any additional bias related to the accuracy of numerical weather simulations.

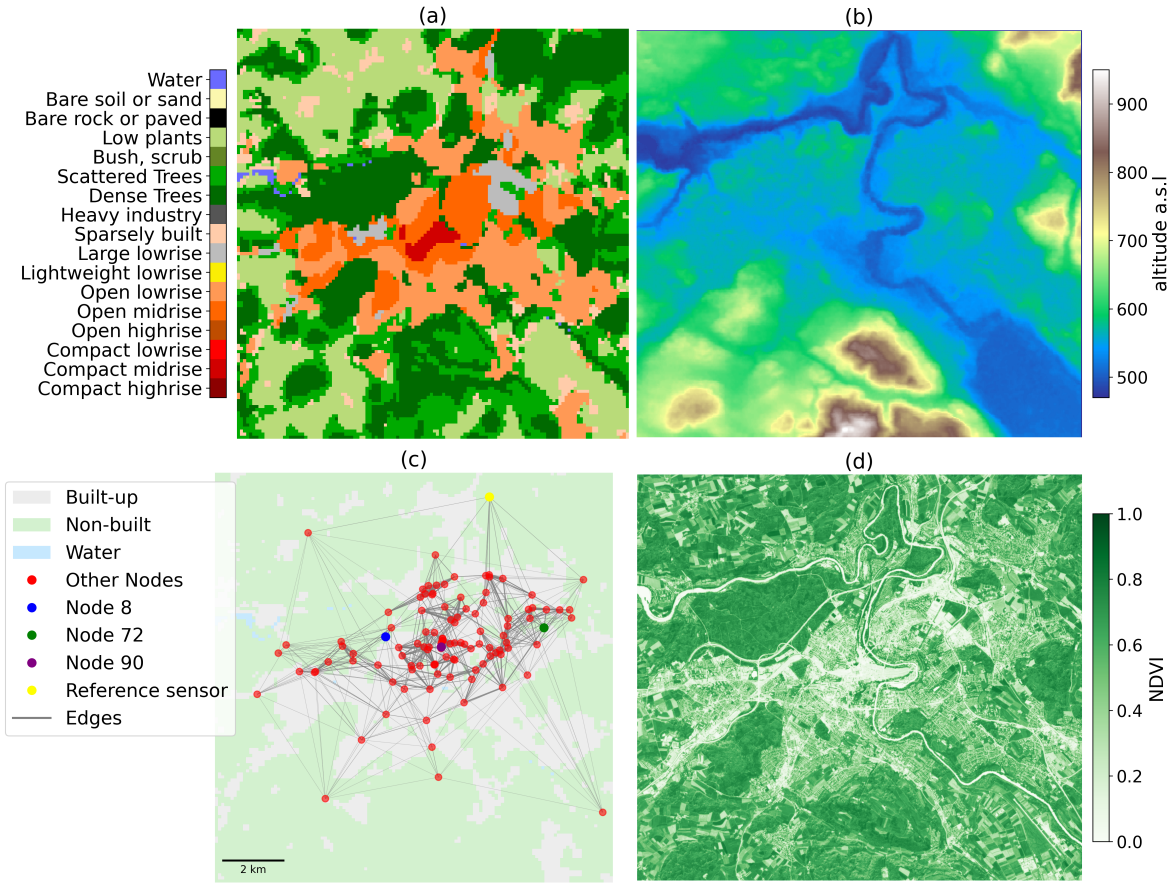


Figure 1: Environmental characteristics of Bern region and location of the sensors, with (a) the Local Climate Zones, (b) elevation, (c) the location of the sensors, including the constructed graph in section 2.2.1 and the sensors plotted in Figure 4, and (d) the average annual Normalized Difference Vegetation Index (NDVI).

2.2 The GNN Model

GNN is designed to learn from data structured as graphs, where nodes represent entities and edges represent relationships between them (Fig. 2). We constructed the graph by assigning a spatial neighborhood to each sensor according to specific rules. An encoder-decoder architecture is used to leverage the graph representation of the sensor network and the temporal dynamics of urban air temperature and weather data. The encoder computes a latent representation of the temporal and spatial dynamics of the last 24 hours, and the decoder uses this representation, along with the weather data, to forecast the air temperature at each node for each hour up to a 24-hour horizon. The three modules are illustrated in Fig. 2 and explained in the following.

2.2.1 Graph Construction Process

For the graph construction process, the method chosen was a weighted k-nearest neighbor algorithm based on a composite distance metric that combines both geographical proximity (d_{geo}) and feature similarity (d_{feature}). The feature similarity is defined as the Euclidean distance between the feature vectors of each node. The composite distance between nodes i

and j is defined as:

$$d_{\text{combined}}(i,j) = \alpha d_{\text{feature}}(i,j) + (1-\alpha) d_{\text{geo}}(i,j), \tag{1}$$

where α is a weighting parameter, and both distance components are normalized to [0,1] range. In our configuration, each temperature station is a node of the graph, and we use the proposed distance to create the edges of the graph.

An edge from node i to node j is established ($e_{i,j}$=1) if and only if j is among the k-nearest neighbors of i according to this composite distance. That is,

$$e_{i,j} = \begin{cases} 1 & \text{if } j \in \text{KNN}(i) \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

and,

$$\text{KNN}(i) = \{j \in V \setminus \{i\} \mid \text{rank}_i(j) \leq k\}, \tag{3}$$

where $\text{rank}_i(j)$ is the rank of the node j in the ordered list of all other nodes in $V \setminus \{i\}$ based on their composite distance d_{combined} to node i, with lower ranks indicating closer proximity. The weight given to an edge is the inverse of the composite distance.

The features used to compute d_{feature} are detailed in Section 2.1.1, standardized for numerical features and one-hot encoded for categorical features (local climate zones). The weighting parameter ($\alpha = 0.7$) was selected by minimizing the root mean squared error (RMSE) on the validation set after 10 training epochs. An additional sigmoid normalization is applied to guarantee the weight repartition. The number of k connected neighbors was set to 10. The resulting graph contains 113 nodes and 1130 non-zero weighted directed edges, shown in Figure 1(c), serving as a basis for the graph-based machine learning model. Using the set of temporal features and the constructed sensor graph, urban air temperatures can be predicted in time and space using the architecture described next.

2.2.2 Encoder

To capture the spatio-temporal dependencies of the UHI phenomenon, the model's encoder employs a Diffusion Convolutional Recurrent Neural Network (DCRNN) [50], which integrates spatial diffusion dynamics and temporal recurrence. This architecture encodes sequential temperature and meteorological data into a latent representation that reflects the complex evolution of UHI intensity across space and time.

The spatial dependencies among the nodes are modeled using diffusion convolution [4], a method that simulates random walk-based information propagation across a graph structure. This enables the model to capture the influence of neighboring locations in a directed and weighted fashion, which is especially relevant in urban environments where heat advection can be asymmetric and location-dependent [e.g., 8, 34, 59, 93]. The DCRNN employs a diffusion convolution with $k = 2$ steps, enabling each node to integrate information from its local and extended neighborhood.

In tandem with the spatial modeling, the DCRNN includes a gated recurrent architecture to process temporal sequences. At each time step, the model receives air temperature data and updates the hidden state by incorporating current observations and spatially diffused information from neighboring nodes. This recurrent mechanism enables the model to learn temporal dependencies over multiple hours and to maintain a memory of past dynamics relevant to future air temperature prediction.

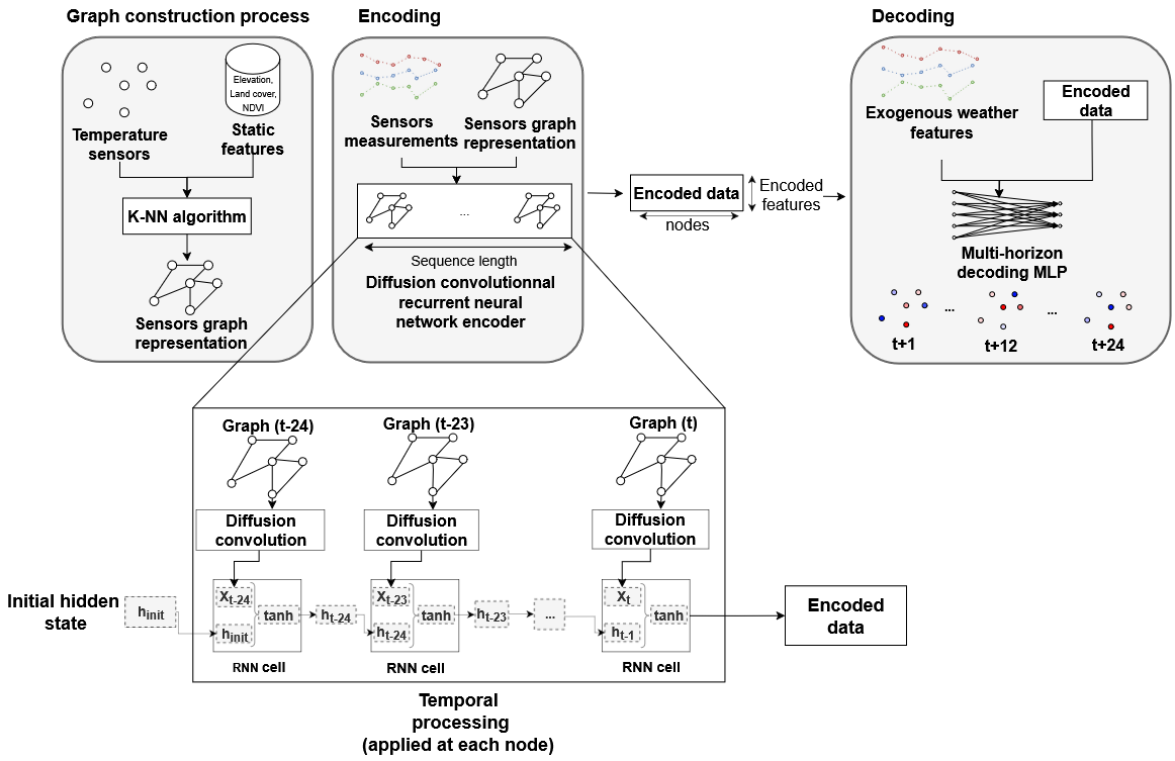


Figure 2: Overview of the DCRNN-MLP architecture for spatio-temporal urban air temperature forecasting. The final output is the air temperature at the different horizons (from 1- to 24-hours ahead) at each sensor.

2.2.3 Decoder

The decoder (Multi-horizon MLP) has been designed to produce multi-step forecasts in parallel, and it draws inspiration from the approach described by Wen et al. [91]. The objective is to incorporate past weather variables to give a global exogenous context for the predictions. The decoder functions in two distinct stages, global and local, each comprising an independent MLP. The global MLP processes the output of the encoder, which is a latent representation of the temporal and spatial dynamics of the temperature field, to create two outputs: a global context vector, shared across all prediction horizons, providing global information on the long-term temperature trend, and a set of horizon-specific vectors that represent time-specific information. The global MLP outputs are concatenated with the exogenous temporal features (weather data). Subsequently, the local MLP generates the actual air temperature forecasts.

2.3 Model Training and Testing

Two different temporal configurations for model training were tested: (i) seasonal, consisting of warm season data only (from May to September) across the years 2019 to 2024; and (ii) all year, covering a continuous period from May 2023 to September 2024. For the seasonal configuration, the model was trained with data from 2019 to 2022, validated using data from 2023, and tested on data from 2024. For the continuous configuration, the dataset was partitioned chronologically into training (70%), validation (10%), and test (20%) sets to preserve temporal dependencies and prevent data leakage. The models were trained on temperature time series with outliers removed using a z-score method (the values above or below 5 standard deviations, i.e., exceeding the 99.99th percentile, were discarded), and for a maximum of

100 epochs, with early stopping implemented to prevent overfitting (training was halted if the validation score did not improve after 10 epochs). To benchmark the performance of the proposed method, a simpler baseline model was used: an RNN [35, 65] applied independently to each sensor, using both the temperature timeseries and the weather data. The RNN processes the timeseries step by step, maintaining a hidden state that summarizes past information and updates it with each new input. The model was trained using the same setup as the GNN (i.e., same train-test split, 100 epochs with early validation) and a hidden layer size of 64.

2.4 Temperature Mapping

The model was then extended to generate high-resolution maps of urban air temperature for each forecast horizon. To convert the sensor-level temperature forecasts into spatially continuous fields, we use a machine-learning-based spatial regression approach relying on XGBoost [15]. This approach explains temperature variability using the full set of auxiliary spatial predictors described in Section 2.1.1. In this way, the method accounts for deterministic effects of environmental and geographical factors (e.g., land use, elevation, NDVI), producing temperature maps at high resolution. The XGBoost model is fine-tuned using cross-validation to avoid overfitting.

We construct a 50 m grid over the Bern study area, where each cell contains the full set of environmental and geographical predictors used for the sensor network (Table A.1). For each prediction horizon, an XGBoost regression model is trained using the sensors' forecasted temperatures as targets and the corresponding predictors as inputs. The trained model is then applied to the full grid to obtain the final high-resolution temperature surface and corresponding UHI-intensity maps.

3 Results and Discussion

3.1 Temperature Forecasting with Past Weather Data

To highlight the benefits of a graph-based approach for temperature forecasting in urban contexts, we start by comparing our model with a simple RNN model. Table 1 describes the average metrics obtained over all the nodes for the DCRNN-MLP, a baseline RNN model and a null model that predicts urban temperatures using only the rural reference temperature.. For the setup using past weather data, The DCRNN-MLP model obtains better metrics for all the horizons and both setups, showing that the GNN approach improves the simulation performance for every forecasting horizon when no future weather forecast is available. In contrast, when future observations are available for training and predictions, the RNN models demonstrate strong performance and consistently outperform the DCRNN-MLP. A notable result is the slightly better performance obtained by the DCRNN-MLP on the continuous dataset compared to the seasonal one, for longer prediction horizons (RMSE of 2.99K versus 3.08K), despite being trained on a smaller amount of data. This is promising for the application of the model to other cities, especially when continuous datasets are available.

Wang et al. [89] showed that, when the horizon is below 4 hours, the models accuracy remains stable, between 1.35 ± 0.60K, while the RMSE was around 3K for the 12- and 24-hour horizons, although the number of studies is limited. Our results indicate slightly lower performance (Table 1), which is expected due to the larger number of sensors used (113 compared to 53 in Yu et al. [97] and 22 in Yu et al. [98]). Furthermore, we forecast air temperature at an hourly resolution (for each hour from 1 to 24), while most existing studies consider only a few discrete forecast lead times (e.g., 1-, 4-, and 24-hour) [89], thus the minimization of the objective function is done at all horizons with no discrimination. Despite

these challenges, the performance gains over the baseline model demonstrate the effectiveness of the DCRNN-MLP in capturing complex spatio-temporal dynamics. In the context of urban air temperature forecasting, RMSE values in the range of 1.5–2.5 K are generally regarded as acceptable, particularly when the prediction horizon extends beyond a few hours [89]. In comparison, measurement uncertainties themselves have been reported to lie between 0.19–0.34 K at night and 0.78–1.17 K during the day [32], suggesting that a portion of the forecast error may be attributable to sensor accuracy.

Setup	Dataset	Model	1 h		6 h		12 h		24 h	
			RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Past Obs.	Continuous	DCRNN-MLP	**1.09**	**0.80**	**2.30**	**1.68**	**2.95**	**2.16**	**2.99**	**2.31**
		RNN	1.43	1.09	2.79	2.15	3.14	2.47	3.17	2.51
	Seasonal	DCRNN-MLP	**1.14**	**0.81**	**2.38**	**1.74**	**2.99**	**2.20**	**3.08**	**2.42**
		RNN	2.05	1.65	2.87	2.25	3.11	2.42	3.29	2.60
Future Obs.	Continuous	DCRNN-MLP	1.47	1.13	1.71	1.32	1.71	1.33	1.68	1.33
		RNN	**0.99**	**0.74**	**1.54**	**1.19**	**1.57**	**1.23**	**1.36**	**1.04**
	Seasonal	DCRNN-MLP	1.68	1.29	2.10	1.61	2.15	1.66	2.16	1.69
		RNN	**1.04**	**0.79**	**2.07**	**1.56**	**1.99**	**1.52**	**2.04**	**1.33**
Against Reference	-	Null Model	1.92	1.47	1.92	1.48	1.92	1.48	1.93	1.48

Table 1: Air temperature prediction performance on the test set at various prediction horizons, categorized by observation setup. Metrics shown are RMSE and MAE (in K). Bold indicates the best performing model within each setup/dataset group. The last row reports a Null Model that predicts urban temperature using only the rural reference temperature.

3.1.1 Spatial Distribution of Forecasting Errors

Figure 3 presents the spatial distribution of forecasting errors. Although the MAE increases substantially from the 1-hour to the 24-hour prediction horizon (0.8-2.3), the standard deviation remains relatively low: 0.08, 0.11, and 0.13 for one, 12-, and 24-hour of lead time, respectively. The consistency metric is calculated as the average of the normalized mean absolute error and the normalized standard deviation of the errors for each horizon, then averaged across horizons for each station and normalized again between 0 and 1. This score reflects the accuracy of a station relative to the others, with lower values indicating lower accuracy. No clear spatial pattern can be observed on the map. The four regression plots illustrate the relationship between selected station spatial features and the consistency score, along with the corresponding fitted linear regression lines. However, the low correlation coefficients (ranging from –0.12 to 0.05) indicate that these features (the other features in the nodes dataset were also tested) do not account for the spatial variability in consistency, suggesting that additional factors are needed to explain the differences in prediction errors across nodes. This highlights a key limitation of the DCRNN-MLP model: its limited interpretability (often encountered when using black-box models [24, 72]).

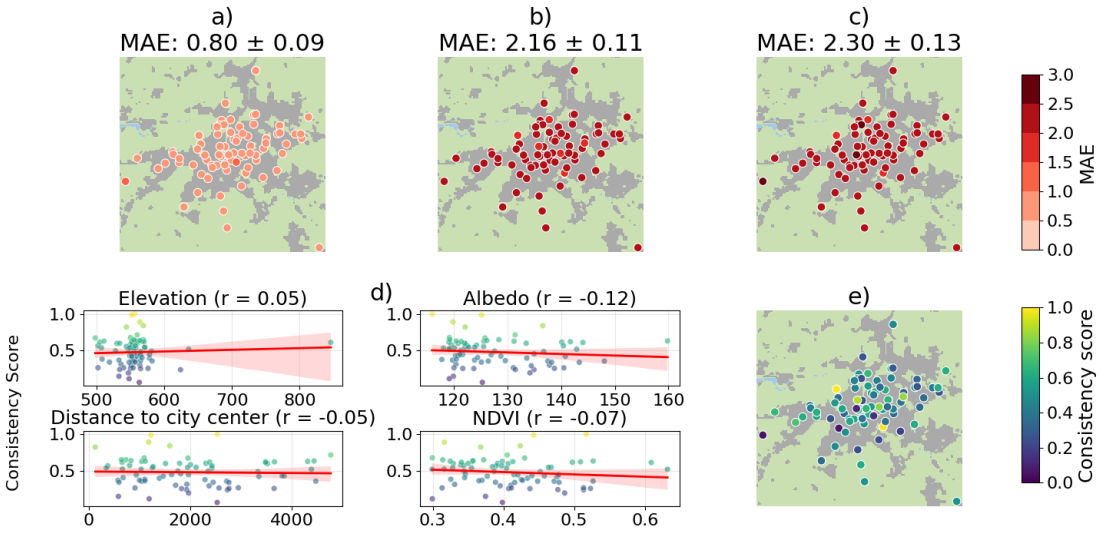


Figure 3: MAE of air temperature forecasting for (a) 1-, (b) 12-, and (c) 24-hour horizons from the DCRNN-MLP (in K). The bottom row illustrates the consistency score of each station forecast errors, averaged across horizons, to quantify the stability and reliability of station-level forecasts. (d) Represents the relation of consistency with four different station features and (e) is the spatial distribution of consistency.

3.1.2 Time Series Forecasting

To gain further insight into the forecasting performance, predicted time series for three representative nodes are shown in Figure 4. Node 8 is located in Bremgartenfriedhof, a cemetery in western Bern acting as a cool island within the urban area; node 90 is in the city center; and node 72 is situated in a residential area in a neighboring municipality of Bern (see locations in Figure 1). The plots compare model forecasts and measured values over three forecasting horizons: 1-, 12-, and 24-hour. At the 1-hour horizon, the model performs well, with RMSE values ranging from 0.903K to 1.108K and MAE values between 0.66K and 0.80K across the three nodes. As the forecast horizon extends to 12 and 24 hours, prediction errors increase, with RMSE rising to 2.26–2.94K (12h) and 2.55–2.84K (24h), and MAE to 1.96–2.26K and 2.01–2.20K, respectively. This performance degradation reflects the growing uncertainty with increasing lead time.

Despite the increased error at 24 hours, the model still captures the overall diurnal temperature trends. However, biases become more evident, particularly in the prediction of daily temperature peaks, which are often under- or overestimated. This is especially noticeable at the horizon 24 hours, where the model tends to overpredict peak temperatures, particularly in the case of an abrupt decrease in temperature. These discrepancies may be linked to the model's reliance on only 24 hours of past weather input data, without any information related to future weather forecasts.

3.1.3 Error Analysis

The detailed analysis of model errors is shown in Figure 5. In Figure 5(a), a prominent yellow band indicates systematically larger errors during the hours between 12:00 and 24:00, especially between 14:00 and 21:00. This effect is most pronounced for forecast horizons exceeding six hours. In contrast, lower errors between 00:00 and 12:00 remain consistent

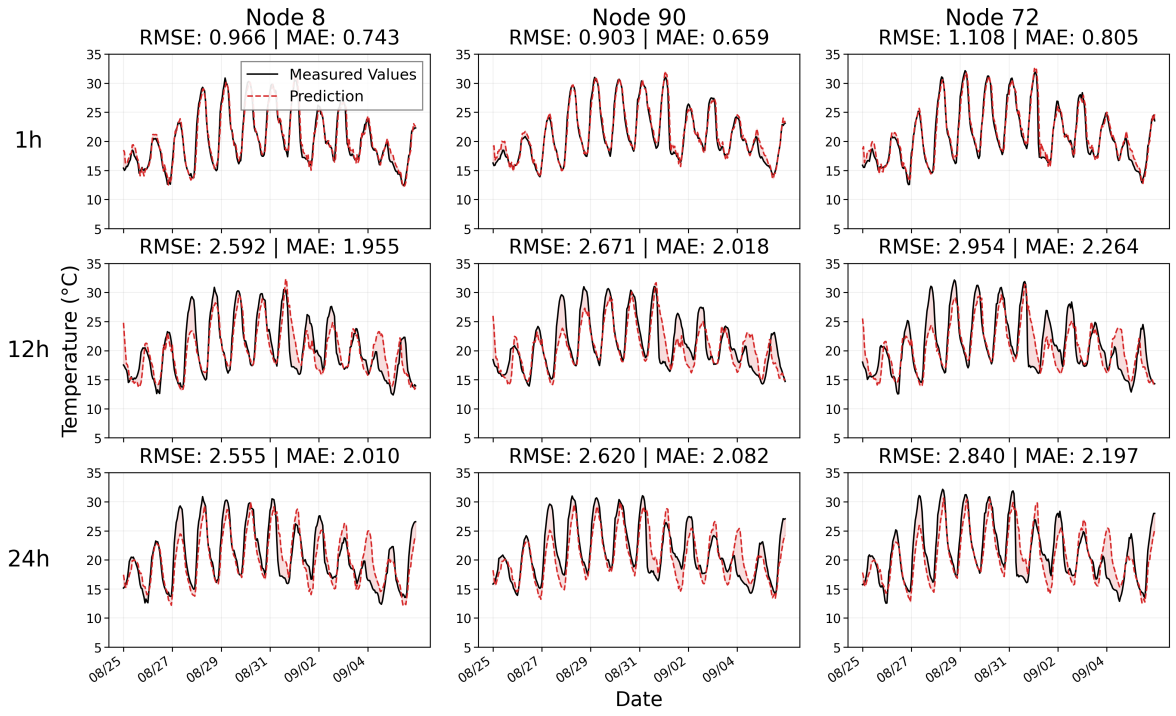


Figure 4: Air temperature forecasting using the DCRNN-MLP at 1, 12, and 24 hours for three different sensors (each row is a different prediction horizon and each column a different node). Sensor 8 is located in a park, sensor 90 is located in the city center, and sensor 72 is located in a residential area. The three sensors illustrated are shown in Figure 1(c).

across all horizons. In Figure 5(b), where blue colors denote underprediction and red denotes overprediction, there are clear overpredictions around 20:00 and 08:00 and underpredictions during the early-morning interval (02:00–04:00) and from 9:00 to 18:00. Panel (c) further shows that MAE increases toward the end of the day and decreases during the early hours while panel (d) demonstrates that RMSE consistently exceeds MAE, particularly at larger error magnitudes, indicating the presence of outlier predictions. Finally, Figure 5 (e) shows the Mean Absolute Percentage Error (MAPE) as a function of the temperature measured at the reference sensor. As expected, MAPE is lowest where sample counts are highest (reference temperatures between 12°C and 25°C).

The error analysis reveals a distinct diurnal bias pattern (Figure 5b), characterized by systematic underprediction of temperatures during daytime hours (approx. 09:00–17:00) and overprediction during the evening and night (approx. 18:00–06:00). The higher RMSE during daytime than nighttime confirms the results by Burger et al. [11], who reported higher uncertainties due to measurement devices during the day. However, during the night, the effect is much more pronounced, resulting in higher prediction accuracy for all horizons. Physically, the observed bias pattern reflects a "damped" thermal response, where the model heats up and cools down too slowly compared to reality, a phase lag that we have demonstrated is largely attributable to the absence of future atmospheric forcing.

3.1.4 Mapping and Kriging Validation

Figure 6 illustrates the observed and forecasted UHI intensity maps for 20:00 UTC on 30 July 2024. The top row displays the spatial distribution of interpolated UHI intensity, showing higher temperatures concentrated in urban local climate zones, particularly in denser areas

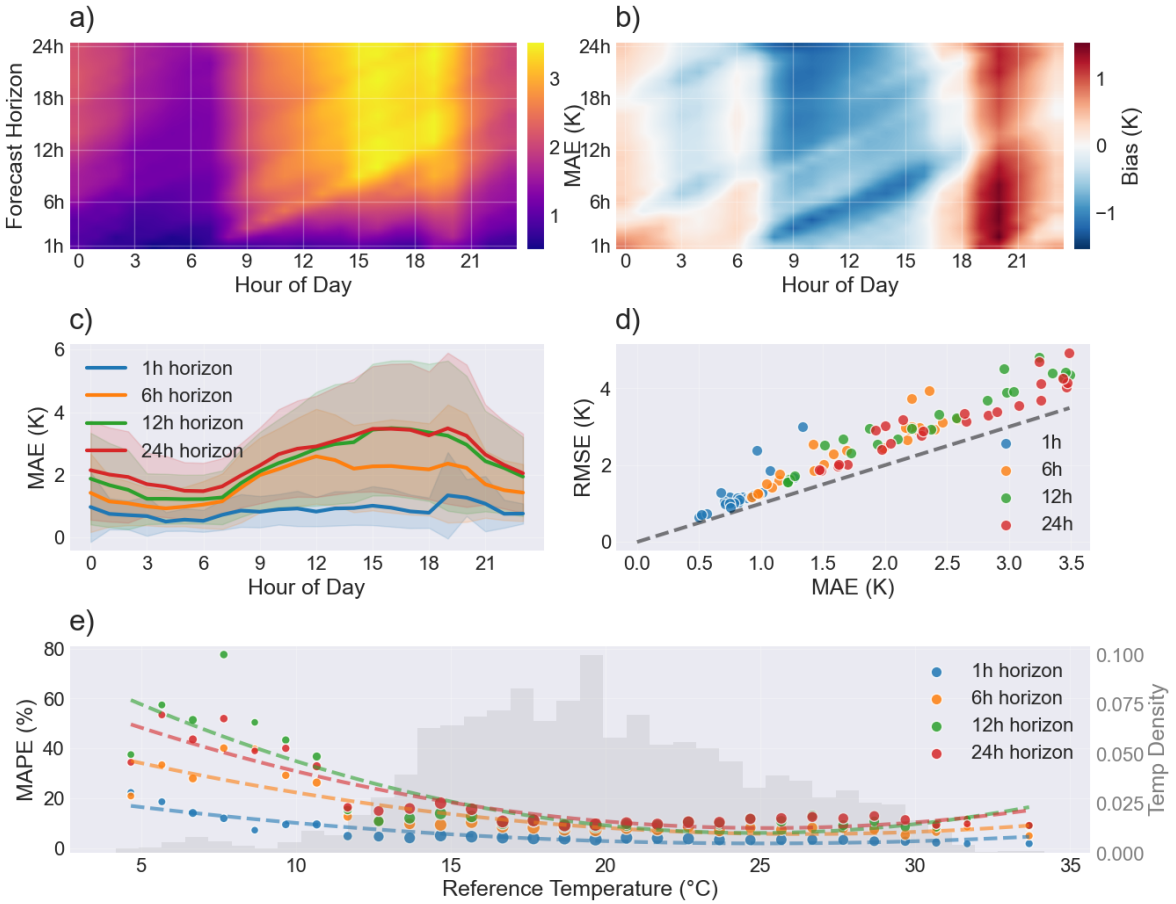


Figure 5: Visualization of model errors and bias across four forecasting horizons (1h, 6h, 12h, and 24h). (a) Mean Absolute Error (MAE) and (b) bias as a function of the target hour of the day (e.g., point "12h, horizon 12" corresponds to the hour 12 when predicted 12 hours ahead). (c) Hourly evolution of the MAE with 95% confidence interval. (d) Relationship between MAE and RMSE, where each point corresponds to a specific hour of the day. (e) Mean Absolute Percentage Error (MAPE) as a function of the temperature measured at the reference sensor.

such as the city center of Bern. In contrast, large vegetated areas like the Bremgartenwald in the northwest of the city act as cooler spots. The UHI values were calculated by subtracting the reference station temperature from the predicted temperature time series, followed by spatial interpolation using XGBoost. The leave one out cross validation (LOOCV) RMSE values, ranging from 0.75 K to 0.85 K, for July 30, indicate good accuracy in the temperature interpolation achieved through this method.

The bottom row of Figure 6 compares predicted and observed UHI values at various station locations for the same time. Forecast accuracy decreases as the time horizon increases, with the RMSE rising from 0.54K at a 1-hour horizon to 0.71 K at a 24-hour horizon, reflecting growing uncertainty in temperature predictions. Despite the increasing uncertainty, the relative ranking of the temperatures is maintained, as shown by the regression coefficients of 0.65 at the 12-hour horizon and 0.67 at the 24-hour horizon. This indicates that, while absolute accuracy decreases, the relative temperature distribution among stations is still captured.

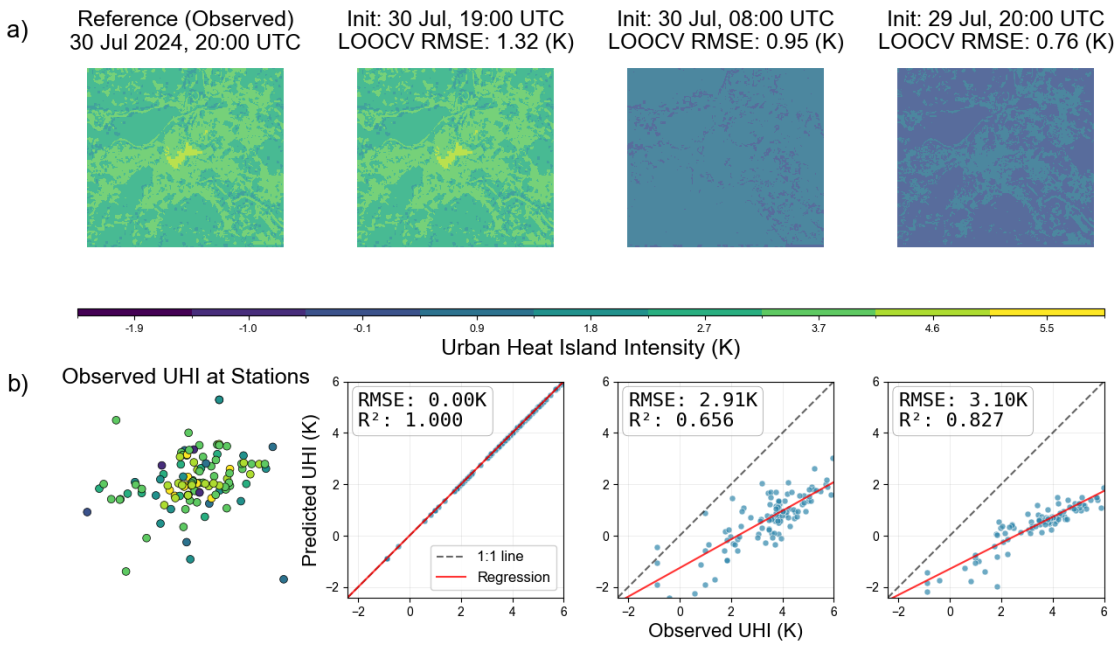


Figure 6: (a) Mapping based on kriging (see Section 2.4) of urban heat island intensity for July 30 at 8PM using forecasting horizons of 1-, 12-, and 24-hour. (b) Observation UHI values and scatter plots representing the distributions of predicted UHI values against measured values at each sensor for July 30, 20:00 UTC, for the different horizon 1h, 12h and 24h (from left to right)

3.2 Impact of Integrating Future Weather Information

To assess the potential benefit of integrating future weather information (e.g., as provided by Numerical Weather Predictions, NWP) into the DCRNN-MLP framework, we conducted an additional experiment using future observed meteorological variables (from the reference station) as a proxy for perfect weather forecasts (i.e., with no model bias). In this setup, the decoder was fed with meteorological data corresponding to the target prediction hours (t+1h to t+24h) rather than the past 24 hours. In addition, compared to the Null model (RMSE = 1.9 K for all forecast horizons), the proposed approach (with continuous data) consistently outperforms across all lead times.

The results of this experiment demonstrates a significant performance improvement, particularly at longer forecasting horizons. The model achieved RMSE values of 1.47 K, 1.71 K, 1.71 K, and 1.68 K for the 1, 6, 12, and 24-hour lead times (Table 1), respectively. Most notably, the integration of future weather data reduced the 24-hour RMSE by 42% compared to the baseline configuration (decreasing from 2.99 K to 1.68 K), stabilizing the error profile over time.

While the DCRNN-MLP benefits significantly from future inputs, the RNN demonstrates even superior performance, Table 1. Because the DCRNN-MLP outperforms the RNN when future inputs are not available, but falls behind once future meteorological variables are provided, the results suggest that the relative performance is strongly influenced by how future exogenous information is injected into the model. In our current DCRNN-MLP design, future variables are used only at the decoder stage, whereas the RNN directly processes these future observations as primary inputs, which likely enables a more direct alignment between

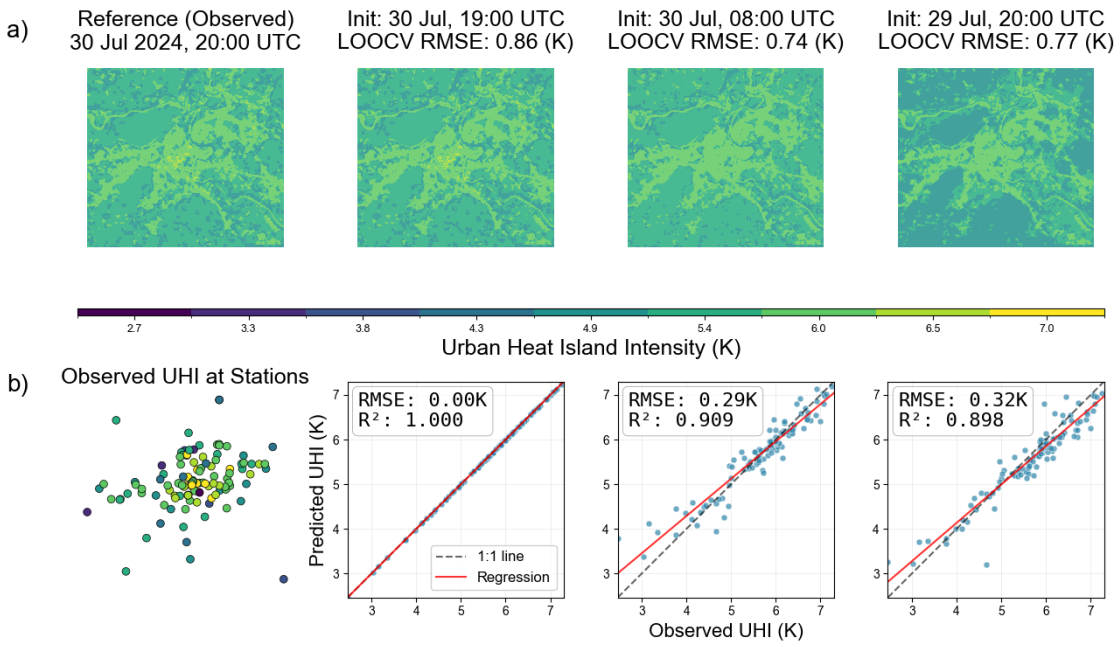


Figure 7: Same as Figure 6 but for the experiment with future weather data.

the meteorological forecasts and the predicted temperatures. This comparison highlights that future meteorological variables can be highly informative, but their benefit depends on the conditioning strategy, and motivates architectures that integrate future information earlier in the spatiotemporal representation learning.

The integration of future atmospheric conditions effectively corrects the temporal lag observed in the DCRNN-MLP with past observations. In the initial configuration (Section 3.1), the model tended to predict temperature peaks with a delay, relying heavily on the previous day's inertia. With future weather data, the phase shift is largely eliminated, and the amplitude of diurnal extremes, both daily maximums and nightly minimums, is captured with greater accuracy. Furthermore, Figure A.1 illustrates the updated time series predictions by the DCRNN-MLP. Contrasting with Figure 4, the phase lag is largely corrected, and peak temperatures are captured with greater accuracy.

3.3 Limitations and Perspectives

In summary, the results here demonstrate that, when using past weather observations, the proposed DCRNN-MLP model outperforms the RNN model used as a baseline across all forecasting horizons and datasets, confirming the benefit of incorporating graph-based spatial dependencies into urban temperature forecasting. While prediction errors increase with longer horizons, the model can capture diurnal temperature dynamics and maintain consistent relative differences between stations, which is important for UHI observations. Yet, the error analysis highlights systematic temporal patterns, with higher errors during the day compared to the night, reflecting the complexity of diurnal urban heat dynamics. These biases carry significant implications for operational deployment. From a public health perspective, the underestimation of daytime peaks is critical, as it could lead to missed early warnings for extreme heat stress, potentially delaying emergency interventions for vulnerable populations. Conversely, the consistent overestimation of nighttime temperatures could trigger false alarms

regarding 'tropical nights', a key metric for human physiological recovery, thereby affecting the credibility of the warning system. For energy sector applications, this bias pattern suggests a potential misalignment in load forecasting: the model may underestimate peak electricity demand for air conditioning during the day while overestimating it in the evening.

Regarding physical limitations, although this study did not explicitly train separate models for different wind regimes, the role of advection is implicitly captured through the inclusion of wind speed and direction as exogenous features in the decoder. Wind-driven heat transport can significantly alter urban microclimate dynamics, particularly by shifting the cooling influence of green spaces or the heating effect of dense built-up areas downwind [45]. While the current graph structure is static, based on a fixed composite distance of geographic proximity and environmental similarity, future iterations of the GNN could incorporate dynamic edge weights. Such an approach would allow the graph connectivity to evolve in real-time, strengthening connections between nodes aligned with the prevailing wind vector and thereby explicitly modeling the directional propagation of heat across the city.

From a methodological perspective, this study employed observed meteorological data as a proxy for future forecasts but it should be noted that operational implementation would rely on coupling the proposed DCRNN-MLP framework with NWP model predictions. In this context, our approach would function as a statistical downscaling tool: it could translate coarse-resolution regional weather forecasts (typically 9–10 km) into hyper-local urban temperature fields (50–100 m). Although operational NWP forecasts introduce their own inherent uncertainty compared to the observed proxies used here, they provide information on all the necessary atmospheric conditions.

A natural next step is to redesign the DCRNN-MLP to incorporate future exogenous variables more directly within the spatiotemporal encoder or the recurrent hidden-state dynamics, rather than restricting them to the decoder input. For example, future meteorological sequences could be used to condition the encoder, or an explicit sequence-conditioning mechanism (e.g., attention-based conditioning). Exploring these alternatives may allow the model to combine the strengths of graph-based spatiotemporal learning with more effective exploitation of meteorological forecasts, and potentially close—or reverse—the performance gap observed when future observations are available.

Finally, while this study relies on a high-density research network unique to Bern, the proposed graph-based framework is inherently adaptable to cities with varying degrees of data availability. Unlike grid-based models (e.g., CNNs) that require uniform data coverage, the graph structure, defined by environmental similarity and geographic proximity, can naturally accommodate the irregular and sparse sensor configurations typical of growing crowd-sourced networks (e.g., Netatmo, Weather Underground). In urban areas with sparser monitoring sensors, the model architecture allows for a shift in reliance: the influence of the local diffusion mechanism (encoder) can be balanced by a stronger weight on the global meteorological context (decoder) and static environmental features (e.g., Local Climate Zones, NDVI). Furthermore, the framework supports transfer learning, where a model pre-trained on a data-rich city like Bern could be fine-tuned for a data-scarce city, leveraging the learned physical embeddings of urban morphological features to predict temperature dynamics in new environments. A key limitation of deploying graph-based models in practice is their limited interpretability, which remains an active area of research [3, 54, 99]. Future work should focus on a comprehensive interpretability analysis, for instance by leveraging explanation frameworks such as GraphLIME [38] or GNNExplainer [95].

4 Conclusion

A spatio-temporal graph machine learning framework is developed for high-resolution urban air temperature forecasting and urban heat island intensity assessment in the city of Bern. By combining a Diffusion Convolutional Recurrent Neural Network encoder with a multi-horizon MLP decoder, the model effectively captures both local diffusion processes across a sensor graph and temporal dynamics driven by past observations and weather conditions. When information on future weather are used in the prediction step, the model performance is further improved. Compared to a baseline RNN applied independently to each node, when no future data is involved, the proposed DCRNN–MLP consistently yields lower RMSE and MAE across 1 h to 24 h horizons, demonstrating its ability to model complex spatio-temporal dependencies. The practical application of the forecasts is further illustrated via regression-kriging interpolation on a 50 m grid, producing high-resolution temperature maps that highlight built environment impacts and cool-island effects within the city of Bern. These maps align with previous observational findings and underscore the utility of the proposed approach for high-resolution heat-risk mapping and warnings. Overall, the DCRNN-MLP model represents a significant step towards accurate, city-scale, and continuous hourly temperature forecasting. Leveraging urban analytics and ML, future model development based on the approach presented here could be used in combination with regional weather forecasting to develop early warning systems for extreme urban heat events.

Acknowledgements

This research was funded by the EPFL/UNIL program Collaborative Research on Science and Society (CROSS) 2025. NP was supported by the Swiss National Science Foundation (SNSF Grant number: 194649, "Rainfall and floods in future cities"). GM acknowledges support from the SNSF Weave/Lead Agency funding scheme (grant number 213995). SF acknowledges support from the Comparative Ecology of Cities project, funded by the Singapore-ETH Future Cities Laboratory Global. The authors would like to thank Stefan Brönnimann as well as MeteoSwiss for providing the meteorological data.

Data Availability

The code required to reproduce the DCRNN-MLP model developed in this study is available in the GitHub repository: https://github.com/urbes-team/DCRNN-MLP

Author Contributions

CR designed the study and performed the analysis under the supervision of GM and MH. GM and NP acquired the funding. All authors interpreted the results, provided feedback that helped shape the analysis, and contributed to writing the manuscript.

A Appendix

Table A.1: Spatial Features

Description	Type	Value Range	Source
Building's height	Numerical	[0,59]	Swiss3DBuildings [21]
Building's count	Numerical	[0, 199]	Swiss3DBuildings [21]
Average distance to the 50, 100 and 200 nearest buildings [m]	Numerical	[0,1001]	Swiss3DBuildings [21]
Population	Numerical	[0, 261]	OFS [27]
NDVI	Numerical	[0.04, 0.79]	Sentinel [20]
Albedo	Numerical	[0.11, 0.20]	MODIS [75]
Elevation [m]	Numerical	[498, 656]	NASA SRTM [63]
Local climate zones	Categorical	[1, 17]	Wudapt [18]
Distance to Bern's city center [m]	Numerical	[0, 7245]	-
Distance to the nearest green space [m]	Numerical	[0, 1131]	SwissTLM3D [67]

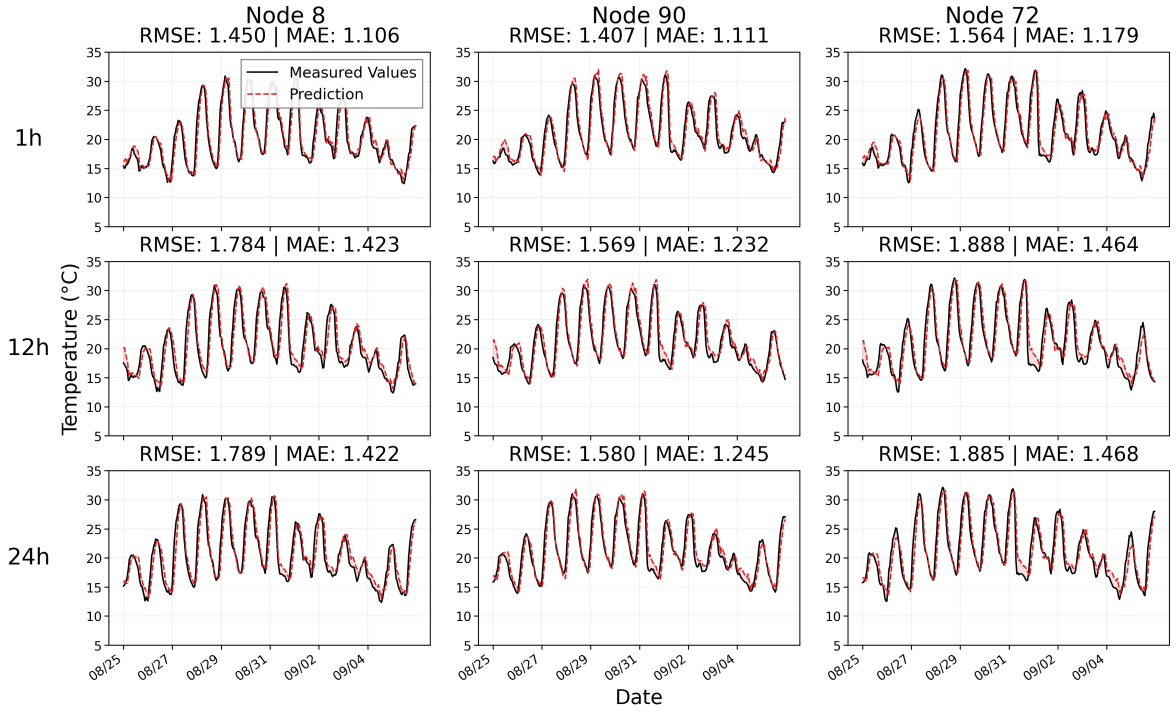


Figure A.1: Air temperature forecasting at 1, 12, and 24 hours for three different sensors for the DCRNN-MLP with future weather observations.

Table A.2: Temporal Features

Description	Type	Value Range	Source
Average hourly vapor pressure [hPa]	Numerical	[5.4, 24.7]	MeteoSwiss [62]
Global irradiation [W/m²]	Numerical	[0, 1034]	MeteoSwiss [62]
Atmospheric pressure [hPa]	Numerical	[935, 967]	MeteoSwiss [62]
Maximum hourly temperature [°C]	Numerical	[2.6, 35.4]	MeteoSwiss [62]
Minimum hourly temperature [°C]	Numerical	[1.9, 34.8]	MeteoSwiss [62]
Hourly precipitations [mm]	Numerical	[0, 23.6]	MeteoSwiss [62]
Average hourly humidity [%]	Numerical	[21.6, 100]	MeteoSwiss [62]
Hourly sunshine duration [h]	Numerical	[0, 1]	MeteoSwiss [62]
Hourly dew point [K]	Numerical	[-1.8, 21.6]	MeteoSwiss [62]
Average hourly wind speed [m/s]	Numerical	[0, 7.9]	MeteoSwiss [62]
Average hourly wind direction [°]	Numerical	[0, 360]	MeteoSwiss [62]
Hour of the day	Numerical	[0, 24]	UNIBE [32]
Month of the year	Categorical	[5, 9]	UNIBE [32]

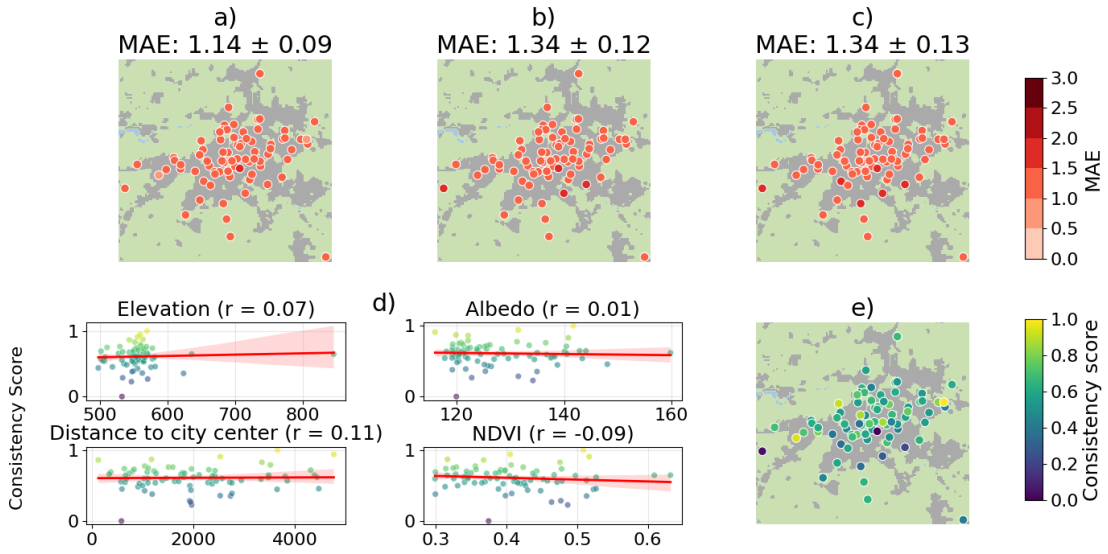


Figure A.2: Results of the regression kriging and the consistency analysis for the DCRNN-MLP with future weather observations.

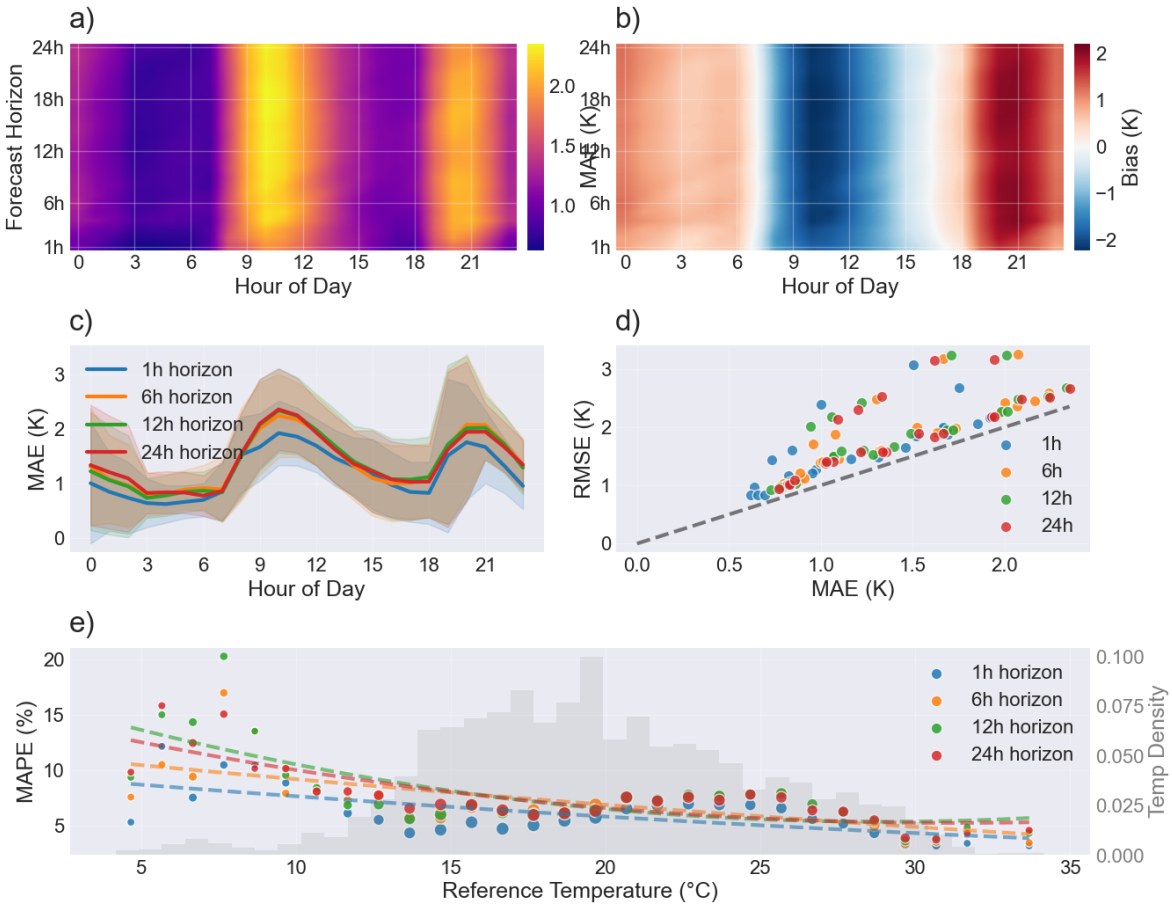


Figure A.3: Global analysis for the DCRNN-MLP with future weather observations.

References

[1] Open data, 2025. URL https://opendatadocs.meteoswiss.ch/fr/.

[2] M. P. Acosta, F. Vahdatikhaki, J. Santos, A. Hammad, and A. G. Dorée. How to bring uhi to the urban planning table? a data-driven modeling approach. *Sustainable Cities and Society*, 71:102948, 2021. doi: 10.1016/j.scs.2021.102948.

[3] M. Altieri, M. Ceci, and R. Corizzo. An end-to-end explainability framework for spatio-temporal predictive modeling. *Machine Learning*, 114(4):114, 2025. doi: 10.1007/s10994-024-06733-6.

[4] J. Atwood and D. Towsley. Diffusion-convolutional neural networks. *Advances in neural information processing systems*, 29, 2016.

[5] A. Badugu, K. Arunab, and A. Mathew. Predicting land surface temperature using data-driven approaches for urban heat island studies: a comparative analysis of correlation with environmental parameters. *Modeling Earth Systems and Environment*, 10(1):1043–1076, 2024. doi: 10.1007/s40808-023-01824-z.

[6] P. Bauer, A. Thorpe, and G. Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015. doi: 10.1038/nature14956.

[7] F. Briegel, J. Wehrle, D. Schindler, and A. Christen. High-resolution multi-scaling of outdoor human thermal comfort and its intra-urban variability based on machine learning. *Geoscientific Model Development*, 17(4):1667–1688, 2024. doi: 10.5194/gmd-17-1667-2024.

[8] O. Brousse, C. Simpson, N. Walker, D. Fenner, F. Meier, J. Taylor, and C. Heaviside. Evidence of horizontal urban heat advection in london using six years of data from a citizen weather station network. *Environmental Research Letters*, 17(4):044041, 2022. doi: 10.1088/1748-9326/ac5ea2.

[9] Bundesamt für Statistik (BFS). Ständige Wohnbevölkerung nach Staatsangehörigkeitskategorie, Geschlecht und Gemeinde, definitive Jahresergebnisse 2021 / Permanent Resident Population by Citizenship Category, Gender, and Municipality, Final Annual Results 2021, Aug. 2022. URL https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung.assetdetail.22504807.html. Published: 25.08.2022, BFS Nummer: su-d-01.02.03.01.01.

[10] M. Burger, M. Gubler, A. Heinimann, and S. Brönnimann. Modelling the spatial pattern of heatwaves in the city of bern using a land use regression approach. *Urban climate*, 38:100885, 2021. doi: 10.1016/j.uclim.2021.100885.

[11] M. Burger, M. Gubler, and S. Brönnimann. Modeling the intra-urban nocturnal summertime air temperature fields at a daily basis in a city with complex topography. *PLoS climate*, 1(12):e0000089, 2022. doi: 10.1371/journal.pclm.0000089.

[12] M. Burger, M. Gubler, and S. Brönnimann. High-resolution dataset of nocturnal air temperatures in bern, switzerland (2007–2022). *Geoscience Data Journal*, 11(4):623–637, 2024. doi: 10.1002/gdj3.208.

[13] X. Cai, J. Yang, Y. Zhang, X. Xiao, and J. C. Xia. Cooling island effect in urban parks from the perspective of internal park landscape. *Humanities and Social Sciences Communications*, 10(1):1–12, 2023. doi: 10.1057/s41599-023-01526-2.

[14] S. Chen, W. Zhang, N. H. Wong, and M. Ignatius. Combining citygml files and data-driven models for microclimate simulations in a tropical city. *Building and Environment*, 185:107314, 2020. doi: 10.1016/j.buildenv.2020.107314.

[15] T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016. doi: 10.1145/2939672.2939785.

[16] C. T. Chiu, K. Wang, A. Paschalis, T. Erfani, N. Peleg, S. Fatichi, N. Theeuwes, and G. Manoli. An analytical approximation of urban heat and dry islands and their impact on convection triggering. *Urban Climate*, 46:101346, 2022. doi: 10.1016/j.uclim.2022.101346.

[17] I. Delgado-Enales, J. Lizundia-Loiola, P. Molina-Costa, and J. Del Ser. A machine learning approach for the efficient estimation of ground-level air temperature in urban areas. *arXiv preprint arXiv:2411.03162*, 2024.

[18] M. Demuzere, J. Kittner, A. Martilli, G. Mills, C. Moede, I. D. Stewart, J. van Vliet, and B. Bechtel. A global map of local climate zones to support earth system modelling and urban-scale environmental science. *Earth System Science Data*, 14(8):3835–3873, 2022. doi: 10.5194/essd-14-3835-2022.

[19] A. El Hachem, J. Seidel, T. O'hara, R. Villalobos Herrera, A. Overeem, R. Uijlenhoet, A. Bárdossy, and L. De Vos. A guide to using three open-source quality control algorithms for rainfall data from personal weather stations. *Hydrology and Earth System Sciences*, 28(20):4715–4731, 2024. doi: 10.5194/hess-28-4715-2024.

[20] E. S. A. (ESA). Copernicus sentinel-2 msi: Multispectral instrument, level-2a, 2023. URL https://scihub.copernicus.eu. Accessed: 2024-12-19.

[21] Federal Office of Topography swisstopo. swiss3Dbuildings 2.0. https://www.swisstopo.admin.ch, 2020. Free geodata and geoservices of swisstopo. Usage requires attribution: ©swisstopo.

[22] G. B. França, V. A. d. Almeida, A. J. d. Lucena, L. d. Faria Peres, H. F. d. Campos Velho, M. V. d. Almeida, G. G. Pimentel, K. d. N. Cardozo, L. B. C. Belém, V. F. V. V. de Miranda, et al. Urban heat island and electrical load estimation using machine learning in metropolitan area of rio de janeiro. *Theoretical and Applied Climatology*, 155(7):5973–5987, 2024. doi: 10.1007/s00704-024-04909-6.

[23] S. Fu, L. Wang, U. Khalil, A. H. Cheema, I. Ullah, B. Aslam, A. Tariq, M. Aslam, and S. S. Alarifi. Prediction of surface urban heat island based on predicted consequences of urban sprawl using deep learning: A way forward for a sustainable environment. *Physics and Chemistry of the Earth, Parts a/b/c*, 135:103682, 2024. doi: 10.1016/j.pce.2024.103682.

[24] M. Garouani, J. Mothe, A. Barhrhouj, and J. Aligon. Investigating the duality of interpretability and explainability in machine learning. In *2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 861–867. IEEE, 2024. doi: 10.1109/ICTAI61863.2024.00130.

[25] J. Gawlikowski, P. Ebel, M. Schmitt, and X. X. Zhu. Explaining the effects of clouds on remote sensing scene classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:9976–9986, 2022. doi: 10.1109/JSTARS.2022.3224090.

[26] L. Ge, K. Wu, Y. Zeng, F. Chang, Y. Wang, and S. Li. Multi-scale spatiotemporal graph convolution network for air quality prediction. *Applied Intelligence*, 51:3491–3505, 2021. doi: 10.1007/s10489-020-02058-z.

[27] B. GEOSTAT. Statistique de la population et des ménages (statpop). 2010. URL https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/geostat.html. Accès: 2024-09-23.

[28] S. Ghorbany, M. Hu, S. Yao, and C. Wang. Towards a sustainable urban future: A comprehensive review of urban heat island research technologies and machine learning approaches. *Sustainability*, 16(11):4609, 2024. doi: 10.3390/su16114609.

[29] M. Giometto, A. Christen, C. Meneveau, J. Fang, M. Krafczyk, and M. Parlange. Spatial characteristics of roughness sublayer mean flow and turbulence over a realistic urban surface. *Boundary-layer meteorology*, 160(3):425–452, 2016. doi: 10.1007/s10546-016-0157-6.

[30] A. Grover and R. B. Singh. Analysis of urban heat island (uhi) in relation to normalized difference vegetation index (ndvi): A comparative study of delhi and mumbai. *Environments*, 2(2):125–138, 2015. doi: 10.3390/environments2020125.

[31] Z. Guan. Urban temperature prediction model based on cnn-bi-lstm model. In *Proceedings of the 7th International Conference on Information Technologies and Electrical Engineering*, pages 53–61, 2024. doi: 10.1109/ICITEE61574.2024.10757757.

[32] M. Gubler, A. Christen, J. Remund, and S. Brönnimann. Evaluation and application of a low-cost measurement network to study intra-urban temperature differences during summer 2018 in bern, switzerland. *Urban climate*, 37:100817, 2021. doi: 10.1016/j.uclim.2021.100817.

[33] W. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

[34] J. Han, X. Zhao, H. Zhang, and Y. Liu. Analyzing the spatial heterogeneity of the built environment and its impact on the urban thermal environment—case study of downtown shanghai. *Sustainability*, 13(20):11302, 2021. doi: 10.3390/su132011302.

[35] J. M. Han, Y. Q. Ang, A. Malkawi, and H. W. Samuelson. Using recurrent neural networks for localized weather prediction with combined use of public airport data and on-site measurements. *Building and Environment*, 192:107601, 2021. doi: 10.1016/j.buildenv.2021.107601.

[36] H. C. Ho, A. Knudby, P. Sirovyak, Y. Xu, M. Hodul, and S. B. Henderson. Mapping maximum urban air temperature on hot summer days. *Remote Sensing of Environment*, 154:38–45, 2014. doi: 10.1016/j.rse.2014.08.012.

[37] L. Howard. *The climate of London: deduced from meteorological observations made in the metropolis and at various places around it*, volume 3. Harvey and Darton, J. and A. Arch, Longman, Hatchard, S. Highley [and] R. Hunter, 1833.

[38] Q. Huang, M. Yamada, Y. Tian, D. Singh, and Y. Chang. Graphlime: Local interpretable model explanations for graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):6968–6972, 2022.

[39] W. T. K. Huang, P. Masselot, E. Bou-Zeid, S. Fatichi, A. Paschalis, T. Sun, A. Gasparrini, and G. Manoli. Economic valuation of temperature-related mortality attributed to urban heat islands in european cities. *Nature communications*, 14(1):7438, 2023. doi: 10.1038/s41467-023-43135-z.

[40] IPCC Core Writing Team, H. Lee, and J. Romero. *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland, 2023. doi: 10.59327/IPCC/AR6-9789291691647. URL https://www.ipcc.ch/report/ar6/syr/.

[41] H. Jiang, Y. Dong, Y. Dong, and J. Wang. Power load forecasting based on spatial–temporal fusion graph convolution network. *Technological Forecasting and Social Change*, 204:123435, 2024. doi: 10.1016/j.techfore.2024.123435.

[42] W. Jiang and J. Luo. Graph neural network for traffic forecasting: A survey. *Expert systems with applications*, 207:117921, 2022. doi: 10.1016/j.eswa.2022.117921.

[43] C. Ketterer and A. Matzarakis. Comparison of different methods for the assessment of the urban heat island in stuttgart, germany. *International journal of biometeorology*, 59:1299–1309, 2015. doi: 10.1007/s00484-014-0940-3.

[44] H. H. Kim. Urban heat island. *International Journal of Remote Sensing*, 13(12):2319–2336, 1992. doi: 10.1080/01431169208904271.

[45] J. Kittner, D. Fenner, M. Demuzere, and B. Bechtel. Analysis of nocturnal urban heat advection using crowd weather stations. *Quarterly Journal of the Royal Meteorological Society*, page e5065, 2025. doi: 10.1002/qj.5065.

[46] M. Kolokotroni, X. Ren, M. Davies, and A. Mavrogianni. London's urban heat island: Impact on current and future energy consumption in office buildings. *Energy and buildings*, 47:302–311, 2012. doi: 10.1016/j.enbuild.2011.12.019.

[47] P. Li and A. Sharma. Hyper-local temperature prediction using detailed urban climate informatics. *Journal of Advances in Modeling Earth Systems*, 16(3):e2023MS003943, 2024. doi: 10.1029/2023MS003943.

[48] Q. Li, E. Bou-Zeid, S. Grimmond, S. Zilitinkevich, and G. Katul. Revisiting the relation between momentum and scalar roughness lengths of urban surfaces. *Quarterly Journal of the Royal Meteorological Society*, 2020. doi: 10.1002/qj.3839.

[49] X. Li, Y. Zhou, S. Yu, G. Jia, H. Li, and W. Li. Urban heat island impacts on building energy consumption: A review of approaches and findings. *Energy*, 174:407–419, 2019. doi: 10.1016/j.energy.2019.02.146.

[50] Y. Li, R. Yu, C. Shahabi, and Y. Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.

[51] Z.-L. Li, H. Wu, S.-B. Duan, W. Zhao, H. Ren, X. Liu, P. Leng, R. Tang, X. Ye, J. Zhu, et al. Satellite remote sensing of global land surface temperature: Definition, methods, products, and applications. *Reviews of Geophysics*, 61(1), 2023. doi: 10.1029/2022RG000778.

[52] W. Lin, D. Wu, and B. Boulet. Spatial-temporal residential short-term load forecasting via graph neural networks. *IEEE Transactions on Smart Grid*, 12(6):5373–5384, 2021. doi: 10.1109/TSG.2021.3093148.

[53] M. J. Lipson, S. Grimmond, M. Best, G. Abramowitz, A. Coutts, N. Tapper, J.-J. Baik, M. Beyers, L. Blunn, S. Boussetta, et al. Evaluation of 30 urban land surface models in the urban-plumber project: Phase 1 results. *Quarterly Journal of the Royal Meteorological Society*, 150(758):126–169, 2024. doi: 10.1002/qj.4612.

[54] N. Liu, Q. Feng, and X. Hu. Interpretability in graph neural networks. In *Graph neural networks: foundations, frontiers, and applications*, pages 121–147. Springer, 2022.

[55] Z. Liu, W. Zhan, B. Bechtel, J. Voogt, J. Lai, T. Chakraborty, Z.-H. Wang, M. Li, F. Huang, and X. Lee. Surface warming in global cities is substantially more rapid than in rural background areas. *Communications Earth & Environment*, 3(1):219, 2022. doi: 10.1038/s43247-022-00539-x.

[56] M. Llaguno-Munitxa and E. Bou-Zeid. Shaping buildings to promote street ventilation: A large-eddy simulation study. *Urban climate*, 26:76–94, 2018. doi: 10.1016/j.uclim.2018.08.006.

[57] Z. Luo, F. Huang, and H. Liu. Pm2. 5 concentration estimation using convolutional neural network and gradient boosting machine. *Journal of Environmental Sciences*, 98: 85–93, 2020. doi: 10.1016/j.jes.2020.05.015.

[58] L. Manickathan, T. Defraeye, J. Allegrini, D. Derome, and J. Carmeliet. Parametric study of the influence of environmental factors and tree properties on the transpirative cooling effect of trees. *Agricultural and forest meteorology*, 248:259–274, 2018. doi: 10.1016/j.agrformet.2017.10.014.

[59] G. Manoli, S. Fatichi, M. Schläpfer, K. Yu, T. W. Crowther, N. Meili, P. Burlando, G. G. Katul, and E. Bou-Zeid. Magnitude of urban heat islands largely explained by climate and population. *Nature*, 573(7772):55–60, 2019. doi: 10.1038/s41586-019-1512-9.

[60] P. A. Mirzaei. Cfd modeling of micro and urban climates: Problems to be solved in the new decade. *Sustainable Cities and Society*, 69:102839, 2021. doi: 10.1016/j.scs.2021.102839.

[61] M. O. Mughal, A. Kubilay, S. Fatichi, N. Meili, J. Carmeliet, P. Edwards, and P. Burlando. Detailed investigation of vegetation effects on microclimate by means of computational fluid dynamics (cfd) in a tropical urban environment. *Urban Climate*, 39: 100939, 2021. doi: 10.1016/j.uclim.2021.100939.

[62] Météo Suisse, 2025. URL https://www.meteosuisse.admin.ch/.

[63] NASA JPL. NASA Shuttle Radar Topography Mission Global 1 arc second. Distributed by NASA EOSDIS Land Processes Distributed Active Archive Center, 2013. URL https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003. Accessed 2024-12-19.

[64] Q. V. Nguyen, J. D. Fernandez, and S. P. Menci. Spatiotemporal graph neural networks in short term load forecasting: Does adding graph structure in consumption data improve predictions? *arXiv preprint arXiv:2502.12175*, 2025.

[65] E. A. Nketiah, L. Chenlong, J. Yingchuan, and S. A. Aram. Recurrent neural network modeling of multivariate time series and its application in temperature forecasting. *Plos one*, 18(5):e0285713, 2023. doi: 10.1371/journal.pone.0285713.

[66] M. Nunez and T. R. Oke. The energy balance of an urban canyon. *Journal of Applied Meteorology and Climatology*, 16(1):11–19, 1977. doi: 10.1175/1520-0450(1977)016⟨0011:TEBOAU⟩2.0.CO;2.

[67] Office fédéral de topographie swisstopo. swisstlm3d, 2024. URL https://www.swisstopo.admin.ch/fr/modele-du-territoire-swisstlm3d.

[68] T. R. Oke, G. Mills, A. Christen, and J. A. Voogt. *Urban climates*. Cambridge university press, 2017. doi: 10.1017/9781139016476.

[69] B. Pioppi, I. Pigliautile, and A. L. Pisello. Human-centric microclimate analysis of urban heat island: Wearable sensing and data-driven techniques for identifying mitigation strategies in new york city. *Urban Climate*, 34:100716, 2020. doi: 10.1016/j.uclim.2020.100716.

[70] Y. Qi, Q. Li, H. Karimian, and D. Liu. A hybrid model for spatiotemporal forecasting of pm2. 5 based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, 664:1–10, 2019. doi: 10.1016/j.scitotenv.2019.01.333.

[71] A. M. Rizwan, L. Y. Dennis, et al. A review on the generation, determination and mitigation of urban heat island. *Journal of environmental sciences*, 20(1):120–128, 2008. doi: 10.1016/S1001-0742(08)60019-4.

[72] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019. doi: 10.1038/s42256-019-0048-x.

[73] M. Santamouris, G. Mihalakakou, N. Papanikolaou, and D. Asimakopoulos. A neural network approach for modeling the heat island phenomenon in urban areas during the summer period. *Geophysical Research Letters*, 26(3):337–340, 1999. doi: 10.1029/1998GL900336.

[74] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008. doi: 10.1109/TNN.2008.2005605.

[75] C. Schaaf and Z. Wang. Mcd43a3 modis/terra+aqua brdf/albedo daily l3 global - 500m v006 [data set], 2015. URL https://doi.org/10.5067/MODIS/MCD43A3.006. Accessed 2025-01-29.

[76] S. I. Seneviratne, X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. D. Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, et al. Weather and climate extreme events in a changing climate. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2021. doi: 10.1017/9781009157896.013.

[77] Y. Shi, L. Katzschner, and E. Ng. Modelling the fine-scale spatiotemporal pattern of urban heat island effect using land use regression approach in a megacity. *Science of the Total Environment*, 618:891–904, 2018. doi: 10.1016/j.scitotenv.2017.11.271.

[78] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, J. G. Powers, et al. A description of the advanced research wrf version 3. *NCAR technical note*, 475(125):10–5065, 2008. doi: 10.5065/D68S4MVH.

[79] I. D. Stewart and T. R. Oke. Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, 93(12):1879–1900, 2012. doi: 10.1175/BAMS-D-11-00019.1.

[80] A. Straub, K. Berger, S. Breitner, J. Cyrys, U. Geruschkat, J. Jacobeit, B. Kühlbach, T. Kusch, A. Philipp, A. Schneider, et al. Statistical modelling of spatial patterns of the urban heat island intensity in the urban environment of augsburg, germany. *Urban Climate*, 29:100491, 2019. doi: 10.1016/j.uclim.2019.100491.

[81] X. Ta, Z. Liu, X. Hu, L. Yu, L. Sun, and B. Du. Adaptive spatio-temporal graph neural network for traffic forecasting. *Knowledge-based systems*, 242:108199, 2022. doi: 10.1016/j.knosys.2022.108199.

[82] J. Tan, Y. Zheng, X. Tang, C. Guo, L. Li, G. Song, X. Zhen, D. Yuan, A. J. Kalkstein, F. Li, et al. The urban heat island and its impact on heat waves and human health in shanghai. *International journal of biometeorology*, 54:75–84, 2010. doi: 10.1007/s00484-009-0256-x.

[83] G. Tanoori, A. Soltani, and A. Modiri. Machine learning for urban heat island (uhi) analysis: Predicting land surface temperature (lst) in urban environments. *Urban Climate*, 55:101962, 2024. doi: 10.1016/j.uclim.2024.101962.

[84] The Weather Company. PWS Network Overview. weather underground. https://www.wunderground.com/pws/overview, 2025. Accessed: 2025-06-16.

[85] H. Torelló-Sentelles, G. Villarini, M. Koukoula, and N. Peleg. Impacts of urban dynamics and thermodynamics on convective rainfall across different urban forms. *Urban Climate*, 62:102499, 2025. doi: 10.1016/j.uclim.2025.102499.

[86] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[87] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.

[88] Z. S. Venter, T. Chakraborty, and X. Lee. Crowdsourced air temperatures contrast satellite measures of the urban heat island and its mechanisms. *Science Advances*, 7 (22):eabb9569, 2021. doi: 10.1126/sciadv.abb9569.

[89] H. Wang, J. Yang, G. Chen, C. Ren, and J. Zhang. Machine learning applications on air temperature prediction in the urban canopy layer: A critical review of 2011-2022. *Urban Climate*, 49, MAY 2023. doi: 10.1016/j.uclim.2023.101499.

[90] H. Wang, J. Zhang, and J. Yang. Time series forecasting of pedestrian-level urban air temperature by lstm: Guidance for practitioners. *Urban Climate*, 56:102063, 2024. doi: 10.1016/j.uclim.2024.102063.

[91] R. Wen, K. Torkkola, B. Narayanaswamy, and D. Madeka. A multi-horizon quantile recurrent forecaster. *arXiv preprint arXiv:1711.11053*, 2017.

[92] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32 (1):4–24, 2020. doi: 10.1109/TNNLS.2020.2978386.

[93] D. Xu, Y. Wang, D. Zhou, Y. Wang, Q. Zhang, and Y. Yang. Influences of urban spatial factors on surface urban heat island effect and its spatial heterogeneity: A case study of xi'an. *Building and Environment*, 248:111072, 2024. doi: 10.1016/j.buildenv.2023.111072.

[94] J. Yang, M. Yu, Q. Liu, Y. Li, D. Q. Duffy, and C. Yang. A high spatiotemporal resolution framework for urban temperature prediction using iot data. *Computers & Geosciences*, 159:104991, 2022. doi: 10.1016/j.cageo.2021.104991.

[95] Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems*, 32, 2019.

[96] B. Yu, H. Yin, and Z. Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.

[97] X. Yu, S. Shi, and L. Xu. A spatial–temporal graph attention network approach for air temperature forecasting. *Applied Soft Computing*, 113:107888, 2021. doi: 10.1016/j.asoc.2021.107888.

[98] Y. Yu, P. Li, D. Huang, and A. Sharma. Street-level temperature estimation using graph neural networks: Performance, feature embedding and interpretability. *Urban Climate*, 56:102003, 2024. doi: 10.1016/j.uclim.2024.102003.

[99] H. Yuan, H. Yu, S. Gui, and S. Ji. Explainability in graph neural networks: A taxonomic survey, 2022.

[100] Z. Zhang, A. Paschalis, A. Mijic, N. Meili, G. Manoli, M. van Reeuwijk, and S. Fatichi. A mechanistic assessment of urban heat island intensities and drivers across climates. *Urban Climate*, 44:101215, 2022. doi: 10.1016/j.uclim.2022.101215.

[101] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE transactions on intelligent transportation systems*, 21(9):3848–3858, 2019. doi: 10.1109/TITS.2019.2935142.

[102] M. Zumwald, B. Knüsel, D. N. Bresch, and R. Knutti. Mapping urban temperature using crowd-sensing data and machine learning. *Urban Climate*, 35:100739, 2021. doi: 10.1016/j.uclim.2020.100739.