https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

Intra-city Scale Graph Neural Networks Enhance Short-term Air Temperature Forecasting

Han Wang[1], Jianheng Tang[2], Jize Zhang[1], Jiachuan Yang[1]

[1]Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, 999077, China

[2]Division of Emerging Interdisciplinary Areas, The Hong Kong University of Science and Technology, Hong Kong, 999077, China

Correspondence to: Jiachuan Yang (cejcyang@ust.hk)

Abstract. Air temperature (T_a) has critical implications for various socioeconomic sectors, yet its dynamics are particularly complex in urban areas due to heterogeneous built environments, landscapes, and diverse anthropogenic activities. Physics-based models struggle with intra-city T_a forecasts due to inadequate urban representation and limited spatial resolution. While weather observation networks offer promising alternatives for direct local T_a modeling, an effective framework to leverage these intra-city data remains lacking. Here, we demonstrate that graph neural networks (GNNs) can harness observation network information to refine T_a prediction at individual locations and elucidate underlying mechanisms. Our novel Mix-n-Scale framework with GNNs achieves over 12% improvement in short-term T_a forecasts compared to conventional time-series approaches. Further model evaluation disentangles performance variations with local T_a variability in diverse spatiotemporal contexts, indicating distinct patterns of intra-city heterogeneity across seasonal and diurnal scales. Our findings establish graph-based approaches for leveraging proliferating urban sensor data and advancing understanding of T_a spatiotemporal dynamics in complex urban environments.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

1 Introduction

Air temperature (T_a) is a crucial meteorological variable that profoundly affects various facets of human welfare (Mora et al., 2017; Yuan et al., 2025; Zhang et al., 2023), including health (Tuholske et al., 2021), energy consumption (Perera et al., 2020; Wang et al., 2023a), and carbon emission (Li et al., 2024), to name a few. Its significance is particularly pronounced in urban areas, where 55% of the global population resides (UN Statistics Division, 2023). Rapid urbanization, characterized by extensive modifications in land use and land cover, has significantly altered the surface energy balance and the overlying climate (Arnfield, 2003; Oke et al., 2017). These transformations, in conjunction with the spatial heterogeneity of the built environment, anthropogenic activities and local landscapes, generate highly localized variations in T_a at scales of approximately 100–1000 meters (Stewart and Oke, 2012). The increasing frequency and intensity of anomalous events under changing climate further complicates T_a pattern in urban areas (Gao et al., 2024; Li and Bou-Zeid, 2013). Accurate and timely local-scale T_a forecasting within cities presents great challenges, despite its critical role in urban management systems (Chen et al., 2024). The conventional approach to T_a forecasting primarily relies on numerical weather prediction (NWP) models, which necessitate solving complex governing equations. However, generating high-resolution forecasts using this physics-based approach presents unique challenges due to urban characteristics, scales issues, and computational demand. First, existing NWP models often lack adequate parameterization schemes to represent complex processes within urban environments (Chen et al., 2011; Nogueira et al., 2022; Sharma et al., 2021). The requirement to specify numerous parameters for urban modules also introduces additional data challenges and uncertainties, which hinder their effective implementation (Chen et al., 2011). Second, substantial knowledge gaps persist in convective scale (<5 km resolution) modelling, including the absence of basic dynamical balances under nonhydrostatic formulations and the inherent complexity of resolving turbulent processes (Kendon et al., 2021; Schär et al., 2020; Yano et al., 2018). Third, the high computational demand of running NWP models, particularly when applying ensemble approaches to address forecast uncertainty, impede their feasibility for real-time operational use. These limitations constrain accurate local T_a forecasts within cities.

Deep learning (DL) has emerged as a promising alternative approach for meteorological variables forecasting. These DL models can be primarily grouped into two paradigms: training with products of physics-based models or direct weather observations. The former paradigm typically relies on ECMWF's ERA5 reanalysis datasets to learn relationships between atmospheric states across successive time steps, and has recently achieved overall superior performance to state-of-the-art operational NWP systems (Bi et al., 2023; Lam et al., 2023; Price et al., 2024). However, this modeling paradigm inevitably inherits issues in urban areas, as the models are trained on data with insufficient urban representations and coarse spatial resolution. The latter paradigm utilizes in situ observations from weather stations or sensors and thus enables models to learn from data that authentically reflect local meteorological conditions (Effrosynidis et al., 2023; Wang et al., 2023b). The typical modeling approach adopts a time-series regression framework, wherein sequences of measurements at each individual locations are used to predict their respective values at subsequent time steps (Haque et al., 2021; Salcedo-Sanz et

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

al., 2016; Wang et al., 2024; Yu et al., 2021). However, the forecast accuracy under this framework remains limited and has improved only marginally despite the progressive adoption of increasingly sophisticated DL methods (Elsayed et al., 2021; Wang et al., 2024; Zeng et al., 2022). These limitations may stem from modeling approaches that rely purely on local time-series information, which on the one hand may fail to capture essential spatial contextual information, rendering the learning task underdetermined and semantically ambiguous (Iakovlev & Lähdesmäki, 2024). On the other hand, this inherently overlooks critical interactions with the surrounding environment that may be essential for accurate forecasting. Modeling observational data within cities provides a solution to deliver local-scale T_a forecasts, while its potential remains underexplored.

With the development of graph neural networks (GNNs), which are capable of modeling discrete and irregularly distributed observation sites, pioneering studies have explored their use in connecting observations across locations to leverage spatial information for enhancing meteorological variable forecasting. Most existing efforts have focused on modeling large-scale observational networks sparsely distributed across broad regions, with the primary rationale being to address: 1) the atmospheric transport and advection processes among locations (Wang et al., 2020; Zhou et al., 2022); 2) weather propagation patterns (Wu et al., 2023); and (3) identify certain causal relationships among different cities (Li et al., 2023). Despite advances in understanding and modeling large-scale dynamics and their associated spatial interactions, it remains largely unknown whether observational network modeling approaches (i.e. incorporating spatial information) are effective at smaller intra-city scales. Furthermore, the underlying mechanisms and spatial dependencies that drive performance improvements in such scale remain unclear.

To study potential interactions among intra-city observations, we implement two GNNs with distinct spatial information aggregation mechanisms (directed and undirected) for short-term (1-6 hours) T_a forecasting, using local measurements of T_a and wind vectors across 16 locations in Hong Kong (Fig. 1a). In support of these GNN's implementation, we propose a novel framework Mix-n-Scale, which integrates optimization and ensemble processes to address the challenge in configuring graph topologies, particularly when prior knowledge of intra-city scale interactions is limited. Furthermore, we quantify the spatial information impacts on each location based on the GNN's information passing principle and compare the results with conventional time-series models where each location is modeled independently. This allows us to separate the contribution of intra-city spatial information on model behavior and understand the underlying mechanisms. This study offers critical insights into effective frameworks for modeling local observational data and sensor networks, which is increasingly important as crowd-sourced weather sensors continue to proliferate within urban environments (Chapman and Bell, 2018). The flexibility of this framework also makes it well-suited for adaptation to the modeling of similar environmental variables.

This paper is organized as follows. Section 2 provides details of datasets, problem formulation, DL models and their training framework, and metrics used in this study. In section 3.1, we first present the overall spatial characteristics of intra-city T_a. Section 3.2 presents modeling results for overall performance and extreme values, followed by an analysis disentangling the impact of spatial information on forecasting in Section 3.3. The spatiotemporal dynamics of T_a forecast performance are further analyzed in Section 3.4. Section 4 presents the summary and conclusions.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

2 Data and Methods

2.1 Datasets

Hong Kong, a densely populated coastal city at the southern edge of East Asia, features complex atmospheric circulation patterns due to its hilly terrain, land-sea contrasts, and heterogeneous urban morphology. make Hong Kong an ideal setting to examine a model's ability to capture local heterogeneity and intra-city T_a dynamics. In this study, we use hourly meteorological data from 16 weather stations (Fig./ 1a) operated by the Hong Kong Observatory. Although more stations exist, we limit our selection to sites with both T_a and wind observations to ensure complete records for exploring the potential effects of wind. More specifically, three types of variables are incorporated into the model training. The first type includes local, spatially varying observations, including T_a and wind speed (both U and V components). The second type includes global observations that are spatially uniform across sites, such as solar radiation (direct and diffuse) and mean sea level pressure; details and statistics of these variables are provided in Table 1. Additionally, we include spatial and temporal stamps for each site to represent its spatiotemporal context, but we do not incorporate detailed land use or urban morphology data, as the focus of this study is time-series forecasting rather than spatial prediction (estimating T_a at unmeasured sites for concurrent time periods). The entire dataset is divided into three disjoint subsets for training, validation, and testing. The training set covers four full years from 2016 to 2019, while the validation and test sets use data from 2020 and 2021, respectively, for model tuning and final performance evaluation.

Table 1

Statistics of Variables Used for Model Training and Evaluation

Type	Input variable	Range	Mean	Unit	Abbreviation
Global	Direct solar radiation	[0, 3.64]	0.38	MJ/m^2	-
	Diffuse solar radiation	[0, 2.24]	0.31	MJ/m^2	-
	Mean sea level pressure	[977.8, 1037.3]	1013.0	hPa	-
Local	Zonal wind speed*	[-13.7, 7.3]	-0.2	m/s	U
	Meridional wind speed*	[-13.2, 5.5]	-0.9	m/s	V
	2-m Air temperature	[-0.9, 38.2]	23.4	°C	T_a
Temporal	Hour of day	[0, 23]	-	-	-
	Day of year	[1, 366]	-	-	-
	Month	[1, 12]	-	-	-
Spatial	Latitude	[113.92, 114.42]	114.156	degree	Lat
	Longitude	[22.20, 22.55]	22.529	degree	Lon
	Altitude	[4, 955]	120	m	Alt

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

Note. *Positive value of U and V denote the wind is from the west and south, respectively.

2.2 Problem formulation and DL models

The task of T_a forecasting at multiple locations is framed as a spatiotemporal prediction problem that uses existing observations to estimate the state of each location over several subsequent time steps. This is processed through a two-stage modeling approach. First, we embed the temporal dynamics at each location separately using Long Short-Term Memory (LSTM) networks, which are effective for encoding time-series information (Greff et al., 2017). We also employ LSTM combining with decoder as a benchmark for time-series modeling using purely local information (Fig. 1b). Based on the time-series embeddings for each location, we then use GNNs to aggregate spatial information from irregularly distributed neighboring locations (Fig. 1a and c). The forecast horizon is set to six hours in this study, as longer lead times would require capturing large-scale dynamics that fall outside the scope of our target domain. The details of these two stages are as follows:

Temporal dynamics embedding: Let the input at a historical time step t as $X_t \in \mathbb{R}^{N\times F}$, where N represents the number of nodes (i.e., weather stations) and F denotes the number of predictor features. The LSTM captures the temporal evolution by processing observations from T previous time lags steps, yielding a set of temporal embeddings $\boldsymbol{h} = \{h_1, h_2, \dots, h_N\}$, with each $h_i \in \mathbb{R}^{F'}$ where F' is the dimensionality of the temporal embedding for nodes from 1 to u (Hochreiter and Schmidhuber, 1997). This can be conceptually denoted as:

$$f_{LSTM}(X_t, X_{t+1}, \dots, X_{t+T}) = \boldsymbol{h}. \quad (1)$$

Spatial information aggregating: Let the spatial connections between weather stations as a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ is the set of nodes with their respective temporal embeddings h_i as node features, and $\mathcal{E}$ is the edge denotes the connection between the nodes. Each node $i \in \mathcal{V}$ aggregates the representations from its immediate neighbors, $\{h_u^k, \forall u \in \mathcal{N}(v)\}$, into a single vector $h_{\mathcal{N}(i)}^{k-1}$. The k is the iteration of spatial aggregation (i.e., the depth of the GNN), and $k = 0$ corresponds to the initial embeddings $\boldsymbol{h}$ from the LSTM. We implemented two GNN architectures, GraphSAGE (GSAGE; Hamilton et al., 2017) and graph attention network (GAT; Brody et al., 2021), because they representing two distinct learning mechanism for spatial information. GSAGE employs an undirected graph structure where all neighboring nodes are weighted equally. In contrast, GAT learns directional influences by implementing an asymmetric attention mechanism that dynamically computes neighbor weights (Brody et al., 2021; Veličković et al., 2018), potentially identifying causal relationships and propagation patterns. The details of two models are described mathematically in Text 1 in supporting information S1.

2.3 Mix-n-Scale framework

Although GNNs offer a flexible modeling paradigm for integrating discrete local observations, determining appropriate graph structure remains an open and challenging problem. Specifically, defining appropriate connectivity patterns between

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

locations and selecting the optimal number of neighboring nodes represents a significant challenge. Such graph topologies are typically constructed through trial and error, involving extensive manual experimentation and iterative testing (Chen and Wu, 2022; Ma et al., 2023; Zheng et al., 2024).

This study therefore treats graph formation, along with time lag T, as hyperparameters and uses a greedy sequential method to search for and optimize their optimal configuration. Moreover, one novelty of our approach is that we do not simply use the best-tuned model but additionally employ an ensemble-based approach to combine the top 10% of validated models composed of different graph topologies and time lags. We call this training framework Mix-n-Scale, and we refer to the trained model as a "hyper-model." To the best of our knowledge, such an ensemble-based approach using various graph structures for sensor network modeling has not been studied or examined. Our rationale for employing this framework is twofold: (1) the selection of neighboring stations to establish connections and the time-series length potentially incorporates information from different spatiotemporal scales, enriching the representation of existing information; (2) since DL model training accounts for the majority of computational resources in model development process (conventional trial and error or our optimization process), while each inference (i.e., forecast) can be completed within seconds with minimal computational cost compared to the training stage (Goodfellow et al., 2016), our proposed hyper-model approach incurs marginal additional computational overhead in real-world applications while more effectively leveraging the substantial resources already required for model development.

Specifically, we use tree-structured Parzen estimator (Bergstra et al., 2011) based on the its loss on validation set, examining various edge formation strategies (from self-connection to connection across all neighbors) for the graphs, look-back lengths (from 1 to 200 time steps) for the input time-series, and varying model architecture hyperparameters. The selection process can be formulated as follows:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\lambda}) \in \operatorname{argmin} \mathbb{E}_{(\mathbf{x},y)\in\mathcal{D}}\left[\ell\left(f_\theta(\mathbf{X}, y, \boldsymbol{\theta}, \boldsymbol{\lambda})\right)\right], \quad (2)$$

where ℓ represents the mean squared error loss. X and y denote individual features and labels, respectively, that comprise the dataset $\mathcal{D}$. f_θ represents corresponding test DL architecture, where $\boldsymbol{\theta}$ encompasses all the model trainable parameters; $\boldsymbol{\lambda}$ represents hyperparameters determining the graph structure, time lags and a few learning hyperparameters including learning rate and hidden dimensions. $\mathbb{E}_{(\mathbf{x},y)\in\mathcal{D}}[\cdot]$ stands for the expectation with the distribution over $\mathcal{D}$. The search process iterates 100 times and selects the model based on the top 10% (10 out of 100) $\boldsymbol{\lambda}$ hyperparameter settings.

2.4 Metrics

2.4.1 Temperature variability metrics

The daily T_a evolution pattern can be primarily described by two metrics, including the mean daily value and the magnitude of diurnal variation. In this study, we introduce diurnal temperature standard deviation (DTSD) to quantify and characterize the intensity of diurnal T_a fluctuations at each location, serving as an indicator to show local T_a pattern. For location i, the DTSD is defined as:

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

$$\mathrm{DTSD}_{(\mathrm{i})} = \sqrt{\frac{1}{24D}\sum_{j=1}^{D}\sum_{h=1}^{24}\left(T_{a(i,j,h)} - \bar{T}_{a(i,j)}\right)^2}, \quad (3)$$

where $T_{a(i,j,h)}$ is the $T_a$at location I , on day j, at hour h; $\bar{T}_{a(i,j)}$ is the mean daily T_a at location I on day j; D is the Total number of days in the datasets.

2.4.2 Model evaluation metrics

We calculate the root mean squared error (RMSE) and Bias to evaluate model performance. These metrics are calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{NT_h}\sum_{t=1}^{T_h}\sum_{i=1}^{N}\left(\hat{T}_{a_{(i,t)}} - T_{a_{(i,t)}}\right)^2}, \quad (4)$$

$$\mathrm{Bias} = \frac{1}{NT_h}\sum_{t=1}^{T_h}\sum_{i=1}^{N}\left(\hat{T}_{a_{(i,t)}} - T_{a_{(i,t)}}\right), \quad (5)$$

where $\hat{T}_{a_{(i,j)}}$ is the predicted T_a at location i at time t. $T_{a_{(i,t)}}$ is corresponding true T_a. N is the total number of the locations, and T_h is the total number of hourly samples. Here, a positive bias indicates overestimation, and vice versa for a negative bias.

2.4.3 Local oscillation index (LOI)

LOI is a metric that we proposed based on the graph Laplacian (Hamilton et al., 2017) that quantifies the surrounding information inflow to each node. This is utilized to quantify the impact of spatial information from surrounding nodes on local forecasting. Mathematically, let $T_{a_{(i,t)}}$ be the T_a observed at location i at the time t, and $T_{a_{\mathcal{N}(i,t)}}$ as the mean T_a of the neighboring stations of station i at the same time is calculated as:

$$T_{a_{\mathcal{N}(i,t)}} = \frac{1}{N}\sum_{u\in\mathcal{N}(i)} T_{a_{u,t}}, \quad (6)$$

where $\mathcal{N}(i)$ represents the set of neighboring stations to i, and N is the number of neighbors. For each station i, the deviation of T_a from its neighbors at any given time t is $\Delta T_{a_{(i,t)}}$:

$$\Delta T_{a_{(i,t)}} = T_{a_{(i,t)}} - T_{a_{\mathcal{N}(i,t)}}. \quad (7)$$

Based on $\Delta T_{a_{(i,t)}}$, one can calculate the historical normal deviation of one station from its neighbors by averaging the deviations over records across the training period. The historical normal deviations $\overline{\Delta T}_{a_{(i,h,m)}}$ for location i at hour h and month m is calculated as follows:

$$\overline{\Delta T}_{a_{(i,h,m)}} = \frac{1}{|T_{h,m}|}\sum_{t\in T_{h,m}} \Delta T_{a_{(i,t)}}, \quad (8)$$

where $T_{h,m}$ represents the set of all historical time points corresponding to hour h and month m, and $|T_{h,m}|$ is the number of time points. And then the LOI is calculated as follows:

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

$$LOI_{i,t} = \Delta T_{a_{(i,t)}} - \overline{\Delta T}_{a_{(i,h,m)}}. \quad (9)$$

LOI essentially reflects how a node differs from its surroundings while eliminating climatological differences. This primarily captures the effect of the graph processing procedure and helps disentangle the impact of spatial information. Note that LOI is an hourly metric, rather than reflecting daily deviation.

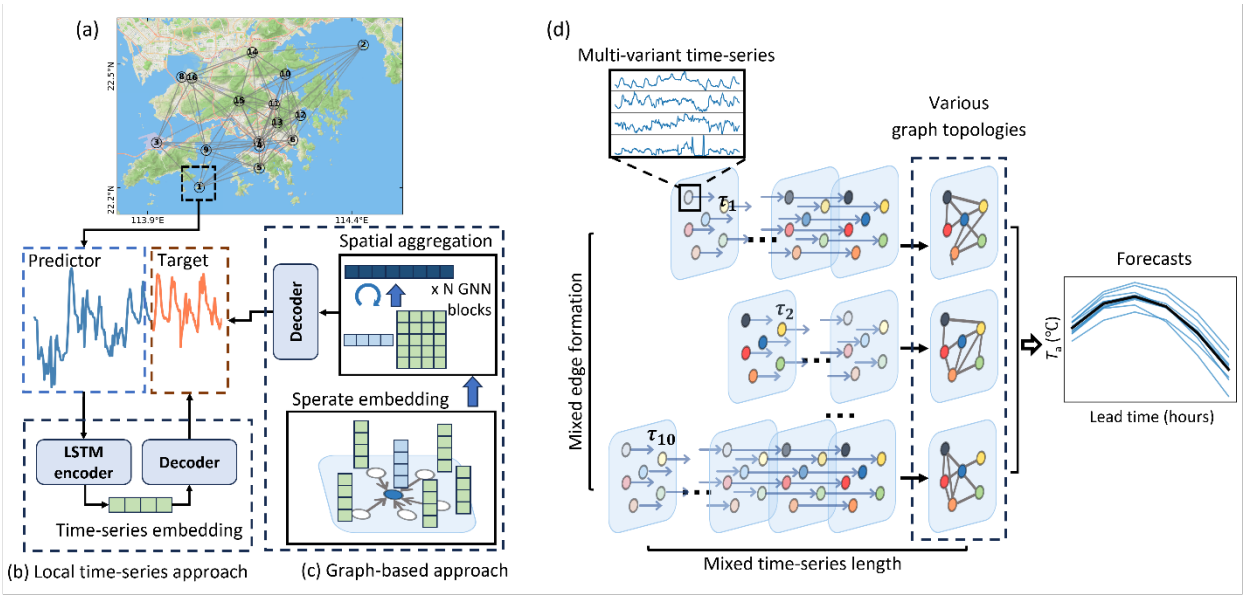

Figure 1. Schematic of the modelling framework. **(a)** Spatial distribution of weather observation stations across Hong Kong (basemap © Mapbox), with location IDs labeled. The edges between stations represent the schematic GNN structure, showing nine connections per node. (b, c) Conceptual diagrams comparing the local time-series modeling approach and the graph-based approach that incorporates spatial information from observation networks **(d)** Overview of the Mix-n-Scale framework, which leverages intra-city observations using diversely configured GNNs.

3 Results and Discussions

3.1 Intra-city T_a characteristics

We first present the intra-city spatiotemporal dynamics of T_a within our study areas. Overall, the mean T_a patterns is relatively homogeneous, with majority sites recording mean values within a narrow range of 23.2°C to 24.2°C. Two notable exceptions are high elevation sites, location 15 (elevation: 955 m) and location 13 (elevation: 572 m), which exhibit the lowest annual mean T_a of 17.6°C and 19.6°C, respectively. In contrast, Hong Kong International Airport, location 3, dominated by concrete structures with high thermal inertia, records the highest mean T_a of 24.8°C.

In comparison, diurnal T_a fluctuation exhibits a more heterogenous pattern. The DTSD (Section 2.4.1) evenly distributed from 1.3°C to 2.5°C, indicating substantial relative spatial variability (Fig. 2b). The lowest DTSD of 1.3°C occurs at the mountain peak (location 15), while the highest value of 2.5°C is observed at location 14 in the northern inland suburban area.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

Notably, diurnal fluctuations tend to be greater in northern areas at shown in the right panel of Fig. 2b, likely due to reduced oceanic thermal moderation and stronger influence from continental air masses (Scheitlin, 2013). The relative magnitude of variation among locations reveals similar mean value patterns but more pronounced differences in diurnal fluctuations.

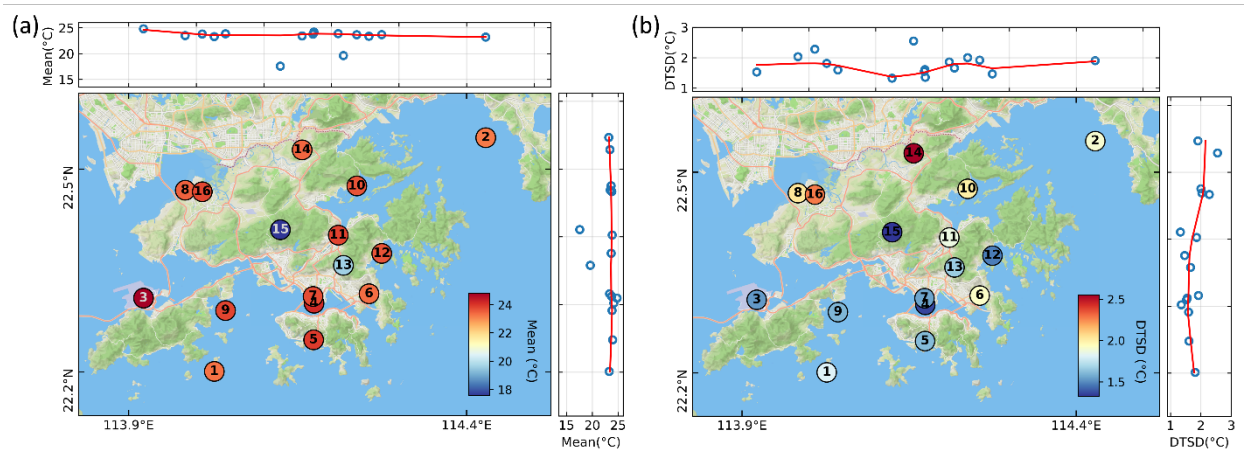

Figure 2. Spatial distribution of **(a)** mean Ta and **(b)** mean diurnal standard deviation (DTSD) over the six year datasets (basemap © Mapbox), with color indicating the magnitude at each location. The upper and right panels show corresponding values along longitude and latitude, respectively, with the solid line indicating a LOESS-smoothed value.

3.2 Overall accuracy of DL models

We evaluate the DL models based on their average performance for 1–6 hour forecasts across 16 weather stations in Hong Kong. The graph-based models consistently outperform approaches that model each local time series individually. The GSAGE achieved the lowest RMSE of 0.96°C, followed by the GAT with 1.03°C, both outperforming the LSTM with 1.06°C. These results highlight the effectiveness of spatial information passed from neighboring stations for local T_a forecasting. For spatial information learning mechanisms, GSAGE's simpler mean aggregation method achieves better accuracy than GAT's adaptive attention approach. While GAT's approach theoretically has higher flexibility by dynamically identifying relationships with neighboring nodes, it may suffer from overfitting problems, despite our inclusion of wind vectors at each location to address potential propagation patterns and our rigorous hyperparameter optimization. This also suggests that clear directed relationships for information propagation from specific "super-nodes" may not exist in T_a forecasting at the intra-city scale. Furthermore, we performed ablation experiments by systematically removing predictors to better understand their contributions. We observe that including wind reduces mean RMSEs from 0.98 to 0.96 °C, while global variables that are uniform across stations (solar radiation and MSLP) failed to further enhance forecast accuracy (Fig. S1 in Supporting Information S1).

Our Mix-n-Scale framework achieves varying performance gains across different DL models (red triangles, Fig. 3a). Since simple LSTM do not involve graph structure, we therefore apply a naïve hyperparameter ensemble that includes models with varying learning rate and hidden dimensions and time lags. However, hyper-LSTM shows only marginal gains over the

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

single LSTM (Fig. 3a). In contrast, the Mix-n-Scale framework produced three times greater improvements when applied to GSAGE compared to LSTM, further indicating the suitability of the framework for graph-based tasks. Overall, Hyper-GSAGE achieves a 12.5% improvement over the best LSTM model, with this superior performance remaining consistent across different ensemble configurations (Fig. S2 in Supporting Information S1). Across all forecast horizons, Hyper-GSAGE consistently outperformed other models while reducing model uncertainty, particularly model spread at longer forecast horizons.

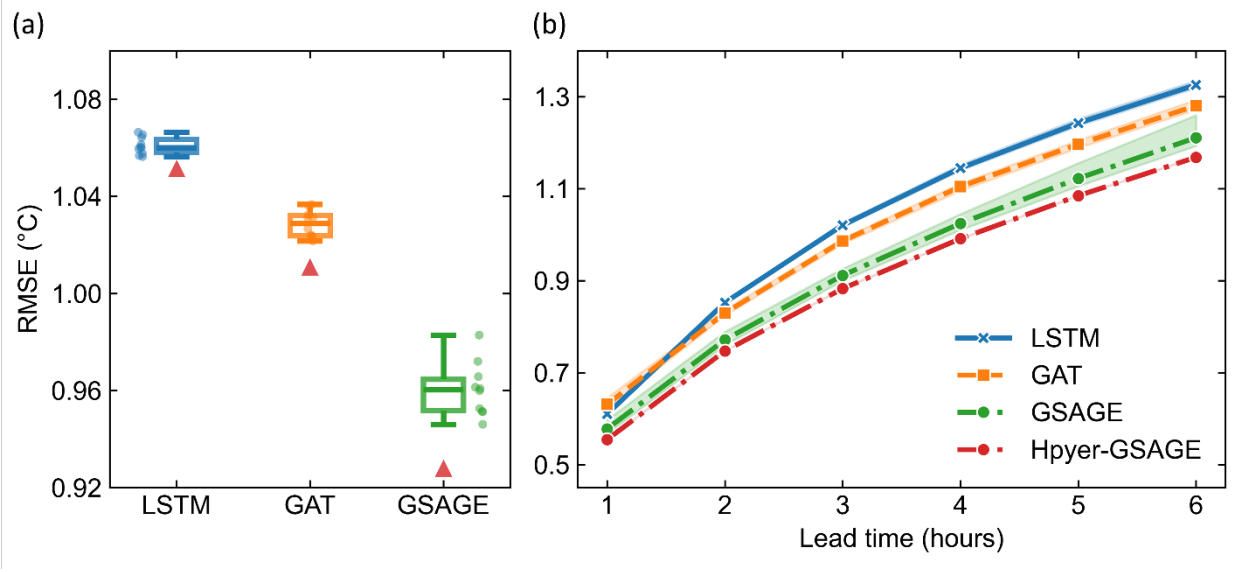

Figure 3. Ta forecast accuracy for the next 6 hours over 16 studied weather stations by different models. **(a)** Overall results of three deep learning models. Each box contains the 10% best individual models (10 out of 100 trained models based on validation results). Box plots show the median (line), 25–75% range (box), and whiskers are drawn to the farthest datapoint within 1.5 inner quantile range. The red triangle denotes the model accuracy with the Mix-n-Scale framework based on the 10% best models. **(b)** Forecast accuracy at different lead times. Shaded areas denote the range of RMSEs among the 10% best models.

Does Hyper-GSAGE preserve extreme values? Given that the model is essentially generated through a multi-model ensemble approach, a major concern is that the results tend to smooth predicted values and sacrifice the ability to capture extreme values (Knutti et al., 2010; Wilks, 2011). Therefore, we examine the distribution of the 5% most extreme values (both warmer and colder) in model forecasts. We find that predicting these values is highly challenging for all models, where we observe rightward-shifted forecasts for colder values and more pronounced leftward shifts for warmer values, reflecting overestimation of low and underestimation of high T_a (Fig. 4a). The greater cold bias for warmer values indicates inherent challenges in capturing extreme high temperatures. However, it is worth noting that Hyper-GSAGE demonstrates better alignment with distribution of observations.

Furthermore, we compare model accuracy extreme conditions based on predicted and corresponding true values (Fig. 4b). For colder values, both GSAGE and Hyper-GSAGE reach comparable results, significantly outperform than LSTM model

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

by reducing RMSE from 1.76°C to ~ 1.50 °C. However, for warmer values, while GSAGE improved RMSE from 1.62°C to 1.53°C, Hyper-GSAGE achieves clear better results (RMSE: 1.41°C) with additional bias reduction from -1.13°C to -0.99°C. These results demonstrate that Hyper-GSAGE enhances performance under both overall and extreme conditions.

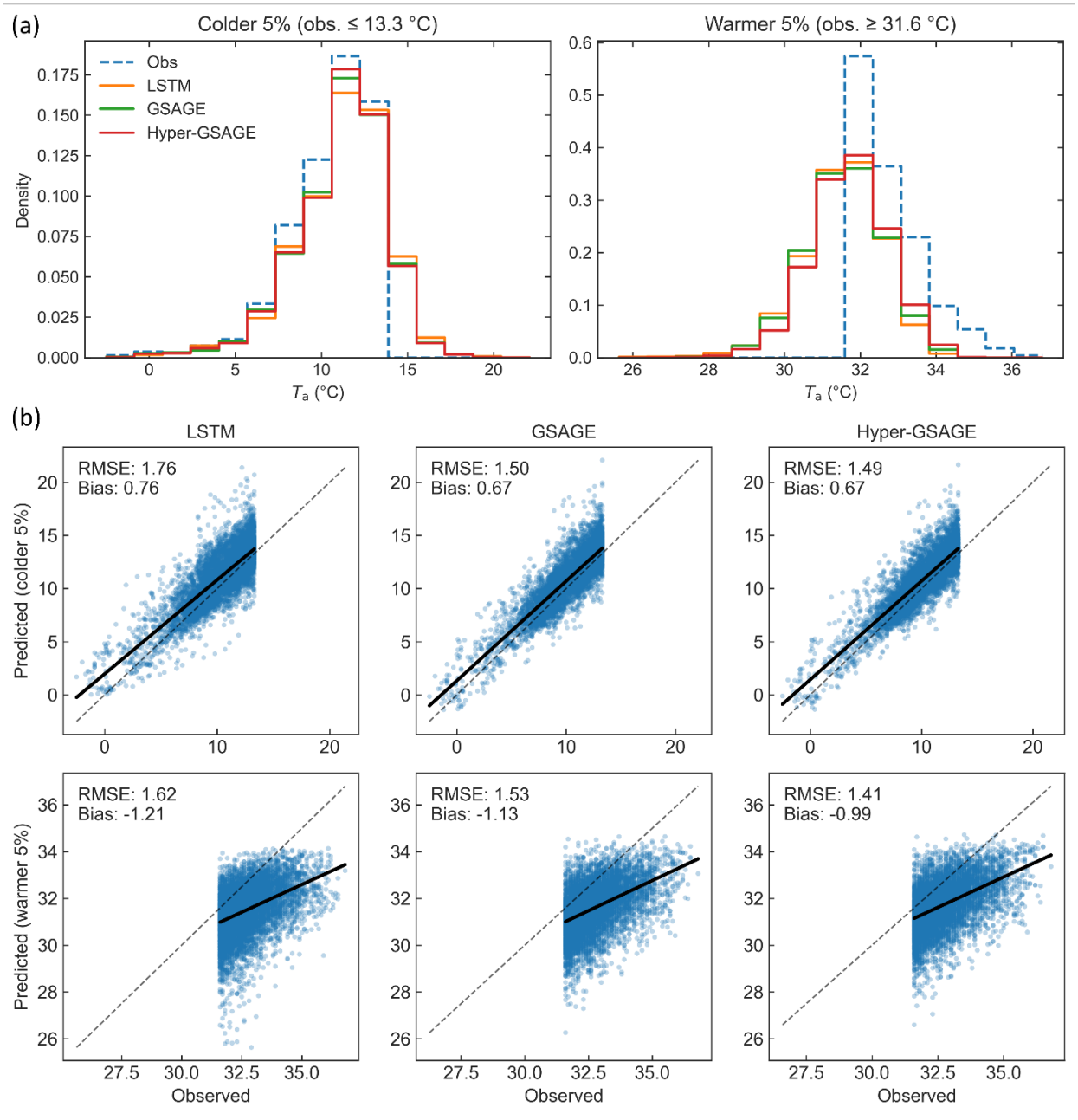

Figure 4. (a) Probability density distributions of the observed T_a and corresponding predictions (6-hour lead time) from the LSTM, GSAGE, and Hyper-GSAGE models for the coldest 5% (obs. ≤ 13.3 °C, left) and warmest 5% (obs. ≥ 31.6 °C, right) of samples. The

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

dashed blue line represents the observed distribution, while solid lines show predictions from each model. **(b)** Scatter plots comparing observed and predicted T_a (6-hour lead time) for the same extremes (coldest 5%, top row; warmest 5%, bottom row). Each column corresponds to a different model: LSTM (left), GSAGE (middle), and Hyper-GSAGE (right). The 1:1 line (dashed) indicates perfect prediction; solid black lines show the linearly fitted regression trend for each case. RMSE and Bias are provided to quantify model performance for the respective extremes.

3.3 Impacts of intra-city scale spatial information

The superior performance of graph-based models demonstrates the critical influence of spatial information, motivating investigation of the underlying mechanisms driving these improvements. This requires quantifying both spatial information inflow to each node and how model behavior changes after incorporating this information. The latter is relatively straightforward to identify by directly calculating the difference between predictions from graph-based models and local LSTM models. However, quantifying spatial information flows to individual nodes is challenging because these flows are learned as high-dimensional latent representations in an end-to-end manner by DL models. To explicitly quantify this information, we propose LOI, an index calculated based on GSAGE's message-passing process (Section 2.4.2) that allows us to track how spatial information influences model behavior. In our context, LOI can be interpreted as the extent to which a location's T_a anomaly deviates from the mean value of its neighboring nodes.

We observe an inverse relationship between the LOI and its impact on T_a forecasts (Hyper-GSAGE minus LSTM, denote as $\Delta\hat{T}_a$ hereafter), as shown in Fig. 5a. This indicates that Hyper-GSAGE tends to adjust a node's prediction upward (positive $\Delta\hat{T}_a$) when its input T_a value is abnormally below its neighbors (negative LOI), as illustrated in Fig. 5b. In other words, this promotes convergence of mean T_a patterns across locations. The rationale behind is that daily mean T_a maintains similar patterns on intra-city scale as noted in Section 3.1, with a limited variance of 0.35°C² among locations (Fig. S3 in Supporting Information S1). The spatially stable mean T_a pattern therefore serves as a dynamic indicator that constrains and refine each node's diurnal amplitude rather than relying solely on local time-series trajectories. Graph regularization naturally enforces such adjustment through its smoothness property (Kipf and Welling, 2017), enhancing model's capacity to modulating local heterogeneous response. We term this effect "mean state regularization" for T_a forecasting.

Fig. 5c presents a case study in location 14 that clearly demonstrates this effect during January 12th-15th when weather starts turning to fine condition (The Weather of January 2021, 2025). The T_a pattern shifts to stronger fluctuation with higher cooling and heating rate. Since this location exhibits cooler T_a than its neighbors, Hyper-GSAGE produces additional upward adjustment in predictions during daytime compared to LSTM, effectively capturing the dynamics, especially for daily peak T_a across those days. However, it is still important to note that LOI provides only a conceptual depiction of relationships that enable us to understand the impact of spatial information. Our interpretation is unable to fully encapsulate the intricacy of DL models that involve propagations through multiple non-linear layers and high-order interactions.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

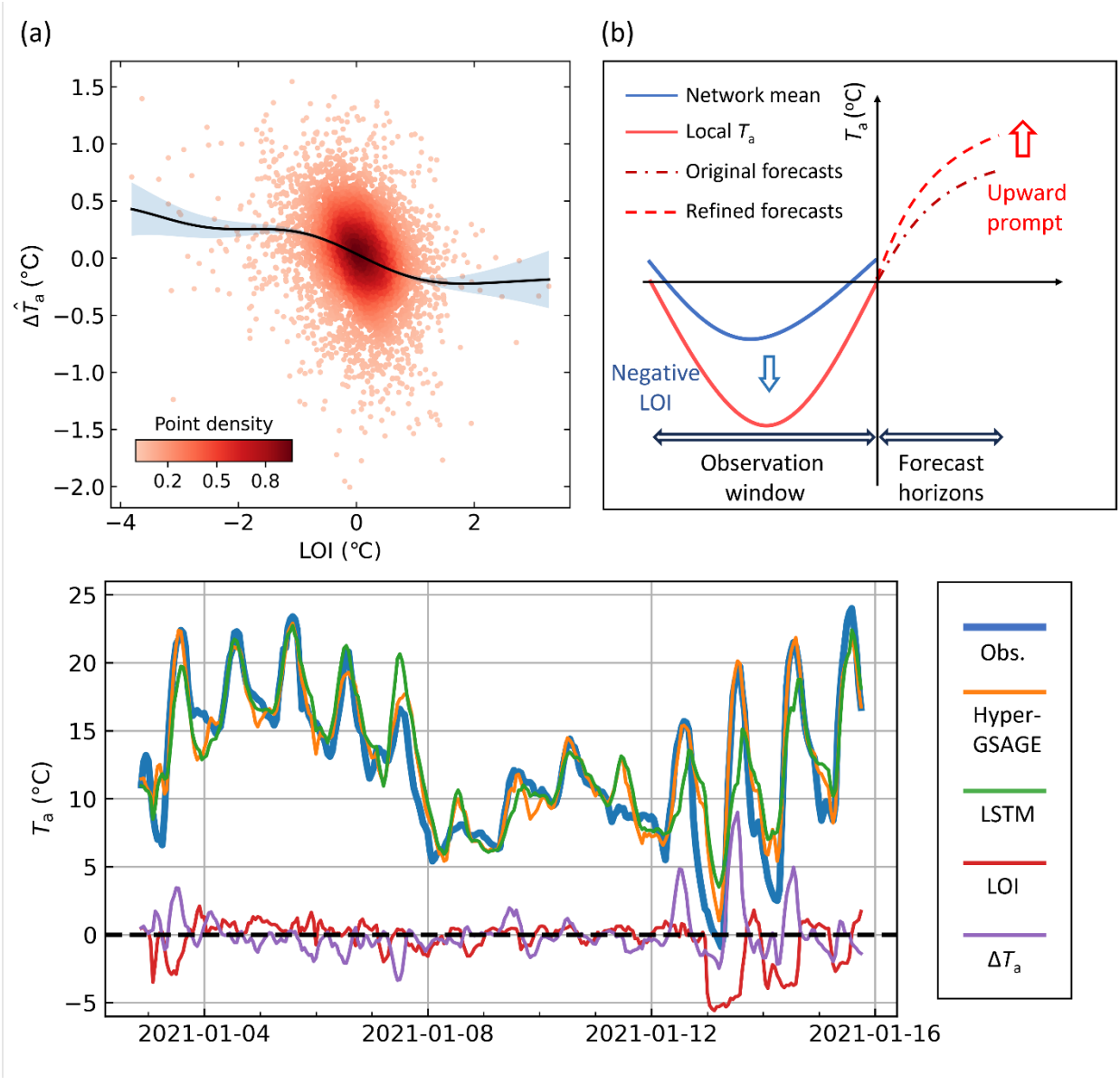

Figure 5. Change in T_a forecast after incorporating spatial information with Hyper-GSAGE. **(a)** Negative relationship between LOI and $\Delta\hat{\boldsymbol{T}}_{\boldsymbol{a}}$. The difference (Δ) is defined as Hyper-GSAGE minus LSTM output at the lead time of 6-hour. Each point in the scatter denotes a daily mean value at a single station, with color indicating the point density. The trend is fitted by Gaussian process regression, with shaded areas denoting the 95% confidence interval of the probabilistic model. **(b)** A diagram illustrating how spatial information influences T_a forecasting at one specific location, where a negative LOI prompts the model to forecast a higher T_a, thereby refining the local magnitude. **(c)** A case study illustrating the temporal evolution of observed T_a and forecasts produced by the LSTM, Hyper-GSAGE models and their difference (ΔT_a) at a 6-hour lead time. The LOI evolution is shown relative to the forecast initialization time to reflect the information can be received by the model.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

3.4 Spatiotemporal dynamics of forecast performance

Following the successful development of our Hyper-GSAGE model, we conducted evaluations of its spatiotemporal performance to better understand forecast characteristics and their variability. The forecast errors demonstrate clear diurnal contrast, with RMSEs increasing during daytime and peaking between 10:00–14:00 (1.60–1.80°C), coinciding with the warmer periods of day (Fig. S4 in Supporting Information S1). In contrast, nighttime forecasts, particularly between 0:00–4:00, show the lowest errors (0.76–0.81°C). This pattern remains consistent when RMSEs are normalized by the mean hourly T_a of the corresponding periods (Fig. S5 in Supporting Information S1). The pronounced diurnal contrast can be primarily attributed to solar radiation-induced perturbations and consequent atmospheric-land interactions, highlighting the inherent challenges in capturing daily peak values. Seasonal error patterns show clear variation, with elevated RMSEs in both winter and summer (mean values of 0.90 and 0.92°C, respectively), peaking in January (0.95°C) and August (0.96°C). Meanwhile, we observe that diurnal profiles vary seasonally, with relatively uniform errors during winter but greater diurnal variations during warmer months (May–September).

Forecast accuracy shows substantial spatial variability, with RMSEs ranging from 0.72 to 1.10 °C (Fig. 6b). Two locations within the most densely developed urban areas show the lowest RMSEs (0.72 and 0.76°C for location 4 and 7, respectively). These locations are surrounded by high-rise buildings in the urban core, where local areas typically have large thermal inertial and reduced ventilation. Location 3, which records the highest mean T_a at the airport, also demonstrates a relatively low RMSE of 0.82°C. In contrast, the highest RMSEs are found at locations 10 and 14 (1.07°C and 1.10°C, respectively), both situated in open areas. As noted in Section 3.1, location 14, situated in suburban northern Hong Kong, exhibits the largest diurnal T_a fluctuation. Location 10 is characterized by a highly complex local thermal environment, being adjacent to both the sea and Hong Kong's largest freshwater reservoir, which likely generates intricate local breeze circulations and greater T_a variability.

The magnitude of local error is highly correlated with the variability of T_a, as measured by the standard deviation (SD) T_a observations at each location. To prevent potential confounding effects arising from temporal variation, we further separated the results into four distinct periods. Across all periods, we observed a significant positive relationship whereby locations with greater T_a variability are associated with higher forecast errors (Fig. 6b). However, this pattern varies dynamically among periods. Summer patterns exhibit distinct diurnal differences. During daytime, local T_a variability diverges significantly across locations (SD from 1.6 to 3.0°C). We treat location 15 as a proxy for background weather conditions, as it is situated atop the city's highest mountain and is therefore minimally perturbed by atmosphere-land interactions. The mountain-top station shows the least local variability and forecast errors, which aligns with Hong Kong's stable summer weather patterns typically dominated by subtropical high-pressure systems. The remaining locations experience greater T_a variability and associated forecast errors during daytime, likely caused by intense solar radiation and subsequent thermal instability and convective turbulence. This variability, along with associated RMSEs, diminishes and converges at night,

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

highlighting the substantial uncertainties and challenges induced by solar radiation in generating T_a instability and spatial heterogeneity during summer. Winter presents a different scenario. Large spread and relatively high variability and RMSEs persist during nighttime, suggesting that non-solar factors predominantly govern winter T_a dynamics. This pattern likely relates to more variable background weather conditions, as evidenced by the mountain peak station (location 15) exhibiting drastically increased variability over this period (Fig. 6b). In this context, local ventilation conditions and thermal inertial likely play crucial roles in regulating local thermal environment exposure to background conditions. We also observed that locations experiencing prevailing northerly winds show elevated forecast errors and T_a variability during winter nighttime (Fig. S6 in Supporting Information S1). Despite this temporal dynamism, the most densely urbanized areas consistently demonstrate lower T_a variability and forecast errors across all periods. These complex spatiotemporal dynamics highlight the diverse driving factors at play, suggesting the need for targeted improvements in model representation and the necessity of period-specific analysis in both intra-city T_a studies and model evaluations.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

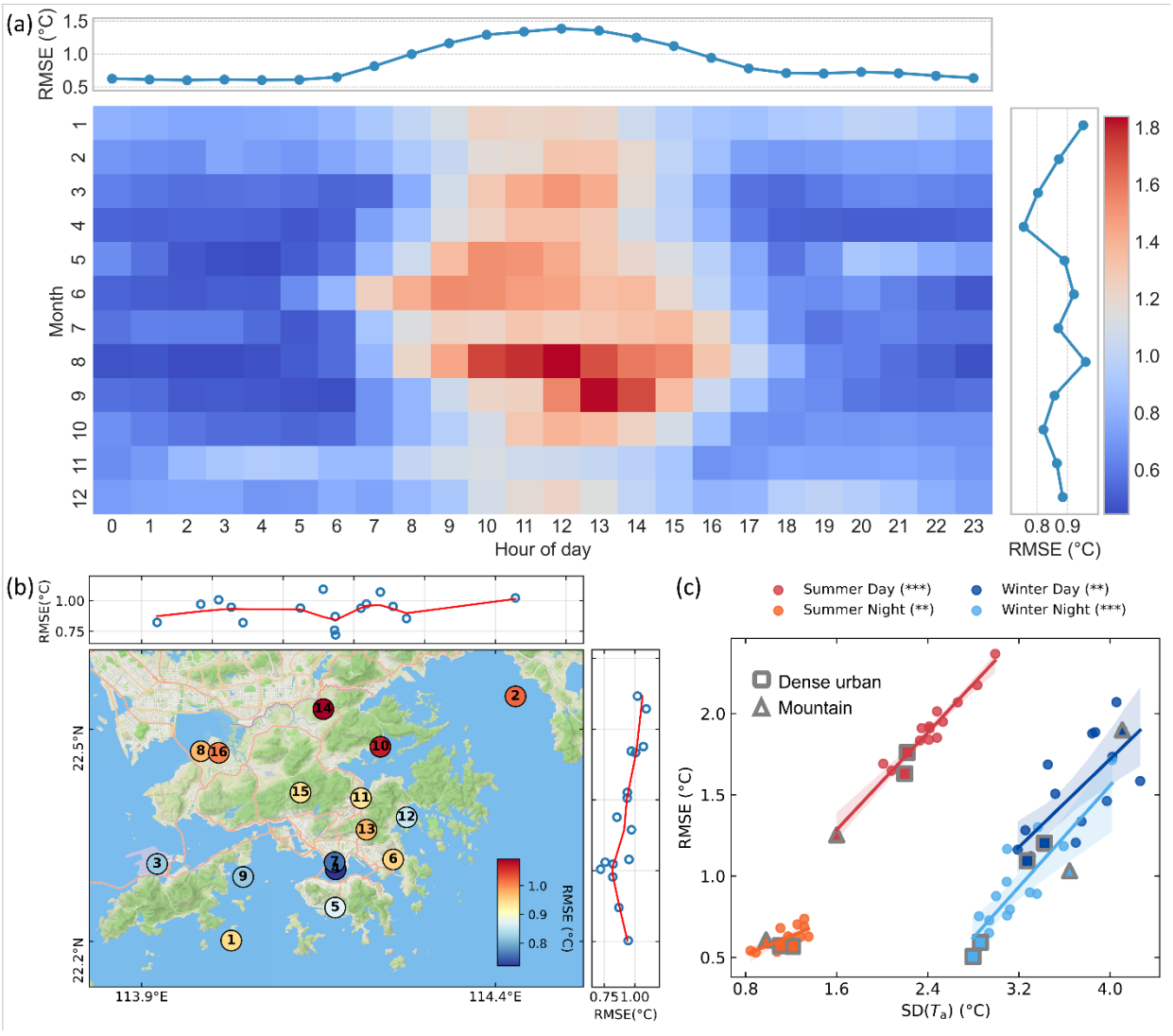

Figure 6. (a) Temporal variation of forecast accuracy. The top and right panels display the mean hourly RMSE aggregated over hours and months, respectively. (b) Same as Fig. 2 but for Spatial distribution of RMSEs (basemap © Mapbox). (c) The relationship between local T_a variability and forecast RMSEs. Each point represents a location, and shaded areas indicate 95% confidence intervals derived from bootstrapping. Dense urban and mountain peak stations are denoted by squares and triangles, respectively. Three and two Asterisks in the legend for each period indicate that the relationship is significant at $p \leq 0.001$ and $p \leq 0.01$, respectively.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

4 Concluding remarks

Our results highlight the necessity of graph-based approaches that connect intra-city observations to improve T_a forecasting at individual locations. This study contributes to a foundational understanding of effective graph formation and the underlying mechanisms through which spatial information enhances forecasts. We demonstrate that an undirected network, learned through a mean state of interaction with neighboring observations, can refine local forecasting by effectively enforcing constraints, a process naturally supported by the GSAGE architecture. With our proposed Mix-n-Scale framework for graph formation and model implementation, the Hyper-GSAGE model produces more accurate forecasts under both overall and extreme conditions and achieving an overall RMSE reduction of over 12.5% for 1–6 hours forecasts compared to the conventional time-series method.

Local T_a exhibits substantial spatial heterogeneity in fluctuation patterns, which strongly correlates with forecast difficulty at corresponding locations. However, the characteristics and drivers of this spatial heterogeneity vary across different temporal periods. Summer exhibits distinct diurnal variations in spatial patterns, suggesting the critical role of solar radiation. In contrast, winter demonstrates more consistent diurnal patterns, where local ventilation and thermal inertia emerge as more critical factors under elevated background T_a variability. Further studies incorporating high-resolution computational fluid dynamics simulations hold great potential for elucidating intra-city airflow dynamics, which could better inform hybrid forecasting models.

Given our focus on intra-city spatial interactions, our models are developed without incorporating meso-scale weather information. We acknowledge this limits the ability to address large-scale weather propagation and therefore constrain the forecast horizon to 6 hours in this study. The current Hyper-GSAGE serves as a foundational framework for modeling local observation networks while retaining the capability to be coupled with large-scale NWP/DL forecast systems for extended forecasts, thereby leveraging advantages from both physics-based and data-driven approaches. With the increasing deployment of IoT weather observation sensors in cities (Chapman and Bell, 2018), such models hold great potential for improving urban management at finer scales, offering a pathway toward more precise and intelligent oversight of urban environmental systems.

4 Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

5 Open Research

The meteorological data for Hong Kong were obtained from the Hong Kong Observatory, which can be acquired from https://www.hko.gov.hk/en/cis/climat.htm. The workflow, model files, and outputs generated during testing and validation are publicly available on Zenodo (Wang, 2025) under the Creative Commons Attribution 4.0 International License.

6 Acknowledgments

This work was supported by the National Natural Science Foundation of China for Excellent Young Scientists (42322903). The authors acknowledge the Hong Kong Observatory's assistance for data preprocessing. We acknowledge the use of OpenAI's ChatGPT as a language refinement tool in the preparation of this manuscript. All content has been thoroughly reviewed and edited by the authors, who take full responsibility for the final publication.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

References

Arnfield, A. J.: Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island, International Journal of Climatology, 23, 1–26, https://doi.org/10.1002/joc.859, 2003.

Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B.: Algorithms for Hyper-Parameter Optimization, in: Advances in Neural Information Processing Systems, 2011.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533–538, https://doi.org/10.1038/s41586-023-06185-3, 2023.

Brody, S., Alon, U., and Yahav, E.: How Attentive are Graph Attention Networks?, in: International Conference on Learning Representations, 2021.

Chapman, L. and Bell, S. J.: High-Resolution Monitoring of Weather Impacts on Infrastructure Networks Using the Internet of Things, https://doi.org/10.1175/BAMS-D-17-0214.1, 2018.

Chen, B., Kong, F., Meadows, M. E., Pan, H., Zhu, A.-X., Chen, L., Yin, H., and Yang, L.: The evolution of social-ecological system interactions and their impact on the urban thermal environment, npj Urban Sustain, 4, 1–11, https://doi.org/10.1038/s42949-024-00141-4, 2024.

Chen, F., Kusaka, H., Bornstein, R., Ching, J., Grimmond, C. S. B., Grossman-Clarke, S., Loridan, T., Manning, K. W., Martilli, A., Miao, S., Sailor, D., Salamanca, F. P., Taha, H., Tewari, M., Wang, X., Wyszogrodzki, A. A., and Zhang, C.: The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems, International Journal of Climatology, 31, 273–288, https://doi.org/10.1002/joc.2158, 2011.

Chen, Y. and Wu, L.: Graph Neural Networks: Graph Structure Learning, in: Graph Neural Networks: Foundations, Frontiers, and Applications, edited by: Wu, L., Cui, P., Pei, J., and Zhao, L., Springer Nature Singapore, Singapore, 297–321, https://doi.org/10.1007/978-981-16-6054-2_14, 2022.

Effrosynidis, D., Spiliotis, E., Sylaios, G., and Arampatzis, A.: Time series and regression methods for univariate environmental forecasting: An empirical evaluation, Science of The Total Environment, 875, 162580, https://doi.org/10.1016/j.scitotenv.2023.162580, 2023.

Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H. S., and Schmidt-Thieme, L.: Do We Really Need Deep Learning Models for Time Series Forecasting?, https://doi.org/10.48550/arXiv.2101.02118, 20 October 2021.

Gao, S., Chen, Y., Chen, D., He, B., Gong, A., Hou, P., Li, K., and Cui, Y.: Urbanization-induced warming amplifies population exposure to compound heatwaves but narrows exposure inequality between global North and South cities, npj Clim Atmos Sci, 7, 1–10, https://doi.org/10.1038/s41612-024-00708-z, 2024.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y.: Deep learning, MIT press Cambridge, 2016.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J.: LSTM: A Search Space Odyssey, IEEE Transactions on Neural Networks and Learning Systems, 28, 2222–2232, https://doi.org/10.1109/TNNLS.2016.2582924, 2017.

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

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

Haque, E., Tabassum, S., and Hossain, E.: A Comparative Analysis of Deep Neural Networks for Hourly Temperature Forecasting, IEEE Access, 9, 160646–160660, https://doi.org/10.1109/ACCESS.2021.3131533, 2021.

The Weather of January 2021: https://www.hko.gov.hk/en/wxinfo/pastwx/mws2021/mws202101.htm, last access: 30 June 2025.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Computation, 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Kendon, E. J., Prein, A. F., Senior, C. A., and Stirling, A.: Challenges and outlook for convection-permitting climate modelling, Philosophical Transactions of the Royal Society A, https://doi.org/10.1098/rsta.2019.0547, 2021.

Kipf, T. N. and Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks, https://doi.org/10.48550/arXiv.1609.02907, 22 February 2017.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in Combining Projections from Multiple Climate Models, https://doi.org/10.1175/2009JCLI3361.1, 2010.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Pritzel, A., Mohamed, S., and Battaglia, P.: GraphCast: Learning skillful medium-range global weather forecasting, https://doi.org/10.48550/arXiv.2212.12794, 4 August 2023.

Li, D. and Bou-Zeid, E.: Synergistic Interactions between Urban Heat Islands and Heat Waves: The Impact in Cities Is Larger than the Sum of Its Parts, https://doi.org/10.1175/JAMC-D-13-02.1, 2013.

Li, P., Yu, Y., Huang, D., Wang, Z.-H., and Sharma, A.: Regional Heatwave Prediction Using Graph Neural Network and Weather Station Data, Geophysical Research Letters, 50, e2023GL103405, https://doi.org/10.1029/2023GL103405, 2023.

Li, P., Wang, Z.-H., and Wang, C.: The potential of urban irrigation for counteracting carbon-climate feedback, Nat Commun, 15, 2437, https://doi.org/10.1038/s41467-024-46826-3, 2024.

Ma, M., Xie, P., Teng, F., Wang, B., Ji, S., Zhang, J., and Li, T.: HiSTGNN: Hierarchical spatio-temporal graph neural network for weather forecasting, Information Sciences, 648, 119580, https://doi.org/10.1016/j.ins.2023.119580, 2023.

Mora, C., Dousset, B., Caldwell, I. R., Powell, F. E., Geronimo, R. C., Bielecki, C. R., Counsell, C. W. W., Dietrich, B. S., Johnston, E. T., Louis, L. V., Lucas, M. P., McKenzie, M. M., Shea, A. G., Tseng, H., Giambelluca, T. W., Leon, L. R., Hawkins, E., and Trauernicht, C.: Global risk of deadly heat, Nature Clim Change, 7, 501–506, https://doi.org/10.1038/nclimate3322, 2017.

Nogueira, M., Hurduc, A., Ermida, S., Lima, D. C. A., Soares, P. M. M., Johannsen, F., and Dutra, E.: Assessment of the Paris urban heat island in ERA5 and offline SURFEX-TEB (v8.1) simulations using the METEOSAT land surface temperature product, Geoscientific Model Development, 15, 5949–5965, https://doi.org/10.5194/gmd-15-5949-2022, 2022.

Oke, T. R., Mills, G., Christen, A., and Voogt, J. A.: Urban Climates, 1st ed., Cambridge University Press, https://doi.org/10.1017/9781139016476, 2017.

Perera, A. T. D., Nik, V. M., Chen, D., Scartezzini, J.-L., and Hong, T.: Quantifying the impacts of climate change and extreme climate events on energy systems, Nat Energy, 5, 150–159, https://doi.org/10.1038/s41560-020-0558-0, 2020.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., Lam, R., and Willson, M.: GenCast: Diffusion-based ensemble forecasting for medium-range weather, https://doi.org/10.48550/arXiv.2312.15796, 1 May 2024.

Salcedo-Sanz, S., Deo, R. C., Carro-Calvo, L., and Saavedra-Moreno, B.: Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms, Theor Appl Climatol, 125, 13–25, https://doi.org/10.1007/s00704-015-1480-4, 2016.

Schär, C., Fuhrer, O., Arteaga, A., Ban, N., Charpilloz, C., Girolamo, S. D., Hentgen, L., Hoefler, T., Lapillonne, X., Leutwyler, D., Osterried, K., Panosetti, D., Rüdisühli, S., Schlemmer, L., Schulthess, T. C., Sprenger, M., Ubbiali, S., and Wernli, H.: Kilometer-Scale Climate Models: Prospects and Challenges, https://doi.org/10.1175/BAMS-D-18-0167.1, 2020.

Scheitlin, K.: The Maritime Influence on Diurnal Temperature Range in the Chesapeake Bay Area, https://doi.org/10.1175/2013EI000546.1, 2013.

Sharma, A., Wuebbles, D. J., and Kotamarthi, R.: The Need for Urban-Resolving Climate Modeling Across Scales, AGU Advances, 2, e2020AV000271, https://doi.org/10.1029/2020AV000271, 2021.

Stewart, I. D. and Oke, T. R.: Local Climate Zones for Urban Temperature Studies, Bulletin of the American Meteorological Society, 93, 1879–1900, https://doi.org/10.1175/BAMS-D-11-00019.1, 2012.

Tuholske, C., Caylor, K., Funk, C., Verdin, A., Sweeney, S., Grace, K., Peterson, P., and Evans, T.: Global urban population exposure to extreme heat, Proceedings of the National Academy of Sciences, 118, e2024792118, https://doi.org/10.1073/pnas.2024792118, 2021.

UN Statistics Division: Goal 11: Make cities and human settlements inclusive, safe, resilient and sustainable., 2023.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y.: Graph Attention Networks, in: International Conference on Learning Representations, 2018.

Wang, C., Song, J., Shi, D., Reyna, J. L., Horsey, H., Feron, S., Zhou, Y., Ouyang, Z., Li, Y., and Jackson, R. B.: Impacts of climate change, population growth, and power sector decarbonization on urban building energy use, Nat Commun, 14, 6434, https://doi.org/10.1038/s41467-023-41458-5, 2023a.

Wang, H.: Scripts for Wang 2025 Intra-city GNN, 2025, https://zenodo.org/records/14536237?preview=1&token=eyJhbGciOiJIUzUxMiJ9.eyJpZCI6ImE2ZGQ5ZmI4LWMxZmUtNGFjMy05MjRhLTAxYjZmZGVjOGQ4YSIsImRhdGEiOnt9LCJyYW5kb20iOiIzZDJiZDYxOGRlZjc2ZWVmZjdmODJmOWU4OGZiOGMxNSJ9.ojJPxvcywT5Q8DrbWNvjl8tFwdLGR9glUqfebY8pWjktusEtkpdD9tY3s0uyr2q5WRHUmu7YRvu4SIu2IeJalg

Wang, H., Yang, J., Chen, G., Ren, C., and Zhang, J.: Machine learning applications on air temperature prediction in the urban canopy layer: A critical review of 2011–2022, Urban Climate, 49, 101499, https://doi.org/10.1016/j.uclim.2023.101499, 2023b.

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

Wang, S., Li, Y., Zhang, J., Meng, Q., Meng, L., and Gao, F.: PM2.5-GNN: A Domain Knowledge Enhanced Graph Neural Network For PM2.5 Forecasting, in: Proceedings of the 28th International Conference on Advances in Geographic Information Systems, arXiv:2002.12898 [cs, eess], 163–166, https://doi.org/10.1145/3397536.3422208, 2020.

https://doi.org/10.5194/egusphere-2025-3429
Preprint. Discussion started: 25 August 2025

Wilks, D. S.: Statistical methods in the atmospheric sciences, Academic press, 2011.

Wu, H., Zhou, H., Long, M., and Wang, J.: Interpretable weather forecasting for worldwide stations with a unified deep model, Nat Mach Intell, 5, 602–611, https://doi.org/10.1038/s42256-023-00667-9, 2023.

Yano, J.-I., Ziemiański, M. Z., Cullen, M., Termonia, P., Onvlee, J., Bengtsson, L., Carrassi, A., Davy, R., Deluca, A., Gray, S. L., Homar, V., Köhler, M., Krichak, S., Michaelides, S., Phillips, V. T. J., Soares, P. M. M., and Wyszogrodzki, A. A.: Scientific Challenges of Convective-Scale Numerical Weather Prediction, https://doi.org/10.1175/BAMS-D-17-0125.1, 2018.

Yu, M., Xu, F., Hu, W., Sun, J., and Cervone, G.: Using Long Short-Term Memory (LSTM) and Internet of Things (IoT) for Localized Surface Temperature Forecasting in an Urban Environment, IEEE Access, 9, 137406–137418, https://doi.org/10.1109/ACCESS.2021.3116809, 2021.

Yuan, Y., Santamouris, M., Xu, D., Geng, X., Li, C., Cheng, W., Su, L., Xiong, P., Fan, Z., Wang, X., and Liao, C.: Surface urban heat island effects intensify more rapidly in lower income countries, npj Urban Sustain, 5, 1–11, https://doi.org/10.1038/s42949-025-00198-9, 2025.

Zeng, A., Chen, M., Zhang, L., and Xu, Q.: Are Transformers Effective for Time Series Forecasting?, https://doi.org/10.48550/arXiv.2205.13504, 17 August 2022.

Zhang, J., You, Q., Ren, G., Ullah, S., Normatov, I., and Chen, D.: Inequality of Global Thermal Comfort Conditions Changes in a Warmer World, Earth's Future, 11, e2022EF003109, https://doi.org/10.1029/2022EF003109, 2023.

Zheng, Y., Zhang, Z., Wang, Z., Li, X., Luan, S., Peng, X., and Chen, L.: Rethinking Structure Learning For Graph Neural Networks, https://doi.org/10.48550/arXiv.2411.07672, 12 November 2024.

Zhou, H., Zhang, F., Du, Z., and Liu, R.: A theory-guided graph networks based PM2.5 forecasting method, Environmental Pollution, 293, 118569, https://doi.org/10.1016/j.envpol.2021.118569, 2022.