Supplementary materials for

Enhancing Local Air Temperature Forecasts with Intra-city Observation Networks and Graph Neural Networks

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

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

Contents of this file

Text S1: Architecture of two graph neural networks (GNNs). They use directed and undirected learning mechanisms to aggregate spatial information from neighboring nodes. The details are as follows:

GSAGE: A mean operator was utilized to aggregate surrounding information; the aggregation can be formulated as:

$$\mathbf{h}_{\mathcal{N}(i)}^{k-1} = \sum_{j \in \mathcal{N}(i)} \mathbf{h}_j^{k-1}, \quad (1)$$

where $\mathbf{h}_j^{k-1}$is the representation of the nodes in node i's immediate neighborhood.

After aggregating neighboring feature vectors, GSAGE concatenates the node's current representation $\mathbf{h}_i^{k-1}$ with aggregated neighborhood vector $\mathbf{h}_{\mathcal{N}(i)}^{k-1}$:

$$\mathbf{h}_i^k = \sigma\left(\boldsymbol{W}^k \cdot \left[\mathbf{h}_i^{k-1} \parallel \mathbf{h}_{\mathcal{N}(i)}^{k-1}\right]\right), \quad (2)$$

where $\boldsymbol{W}$ are learned, and $\parallel$ denotes vector concatenation, and this concatenation can be understood as the simple form of "skip connection". The aggregation process is also illustrated in Figure 1c.

GAT: Contrary to GSAGE, GAT can adeptly assign the importance of their neighbors. We applied this to examine whether there are super nodes that can provide more important information than other nodes. An improved version, GATv2, was applied to avoid static attention problems (Brody et al., 2021) in our study. A scoring function e: $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ assigns the importance score α_{ij} for every edge (j, i) , which indicates the importance of the features of the neighbor j to the node i:

$$\alpha_{ij} = softmax_j\left(e\left(\boldsymbol{h}_i, \boldsymbol{h}_j\right)\right) = \mathrm{softmax_j}\left(\boldsymbol{a}^\top \, \mathrm{LeakyReLU}\left(\boldsymbol{W}^k \cdot \left[\boldsymbol{h}_i \parallel \boldsymbol{h}_j\right]\right)\right), \quad (3)$$

where $\boldsymbol{a}, \boldsymbol{W}$ are learned, and α_{ij} usually is unequal with α_{ji}.

With edge weights, GAT computes the node representation as a weighted average over its neighbors:

$$\boldsymbol{h}_i^k = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} \cdot \boldsymbol{W}\boldsymbol{h}_j^{k-1}\right). \quad (4)$$

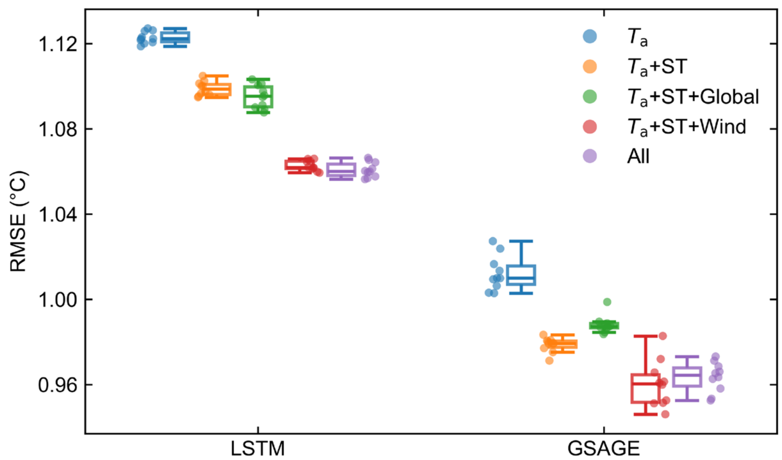

Figure S1. Impact of predictors on Ta forecast accuracy. Performance of LSTM and GSAGE models with various combinations of predictors. ST represent spatial and temporal predictors. Global represents global meteorological predictors (uniform across stations) include direct and diffuse solar radiation, and mean sea level pressure. Wind includes zonal and meridional wind speed at individual stations. Detailed variable descriptions are available in **Table 1**. Box plots show the median (line), 25–75% range (box) based on the 10% best models, and whiskers are drawn to the farthest datapoint within 1.5 inner quantile range.

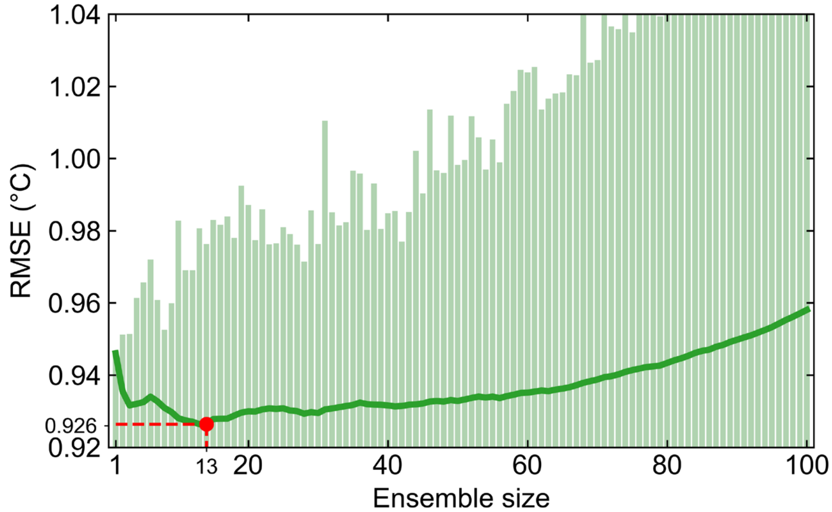

Figure S2. Variation in Hyper-GSAGE performance with different ensemble sizes. Each column represents the performance of an individual GSAGE model, sorted by ascending validation error. The green line denotes the Hyper-GSAGE performance with a corresponding number of best models. It should be noted that the observed increase in RMSE with large ensemble sizes beyond 13 is primarily due to the inclusion of failure models. Conducting additional trials within the optimal hyperparameters range generally achieves a better performance, and this graph is only for illustrative purposes.

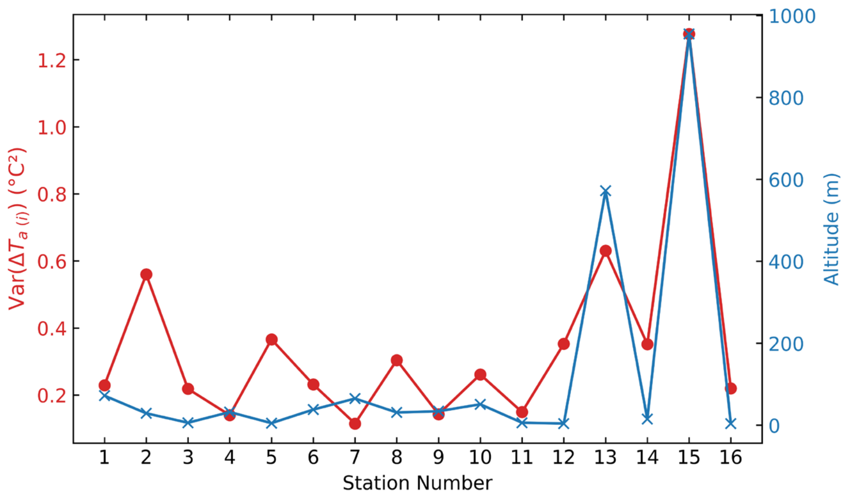

Figure S3. Variance of daily T_a anomalies and altitude at 16 studied stations. The red line shows the variance of daily mean T_a anomalies (left y-axis) for each station, calculated as the deviation of each station's Ta from the mean value across all stations (see equation (7) in Methods) while applied to daily scale. Lower variance indicates more synchronized Ta daily mean evolution between the local station and the global pattern. Blue crosses show the altitude of each station (right y-axis). Station numbers are shown on the x-axis.

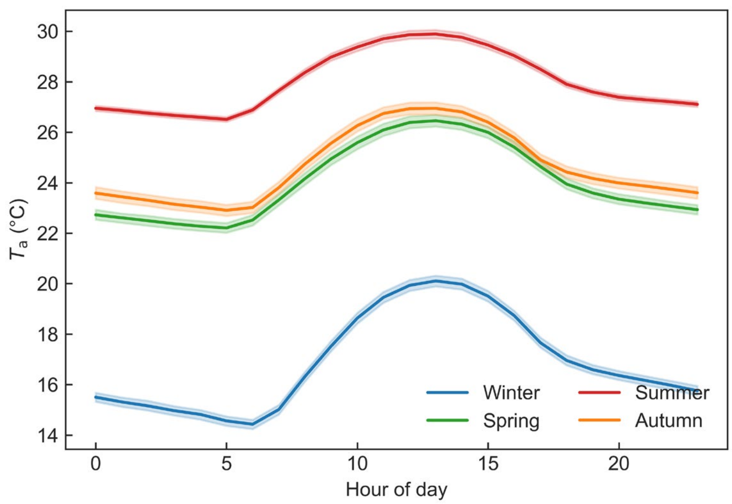

Figure S4. Diurnal variation of T_a cross seasons. Solid lines represent the mean Ta, with shaded bands indicating 95% confidence intervals.

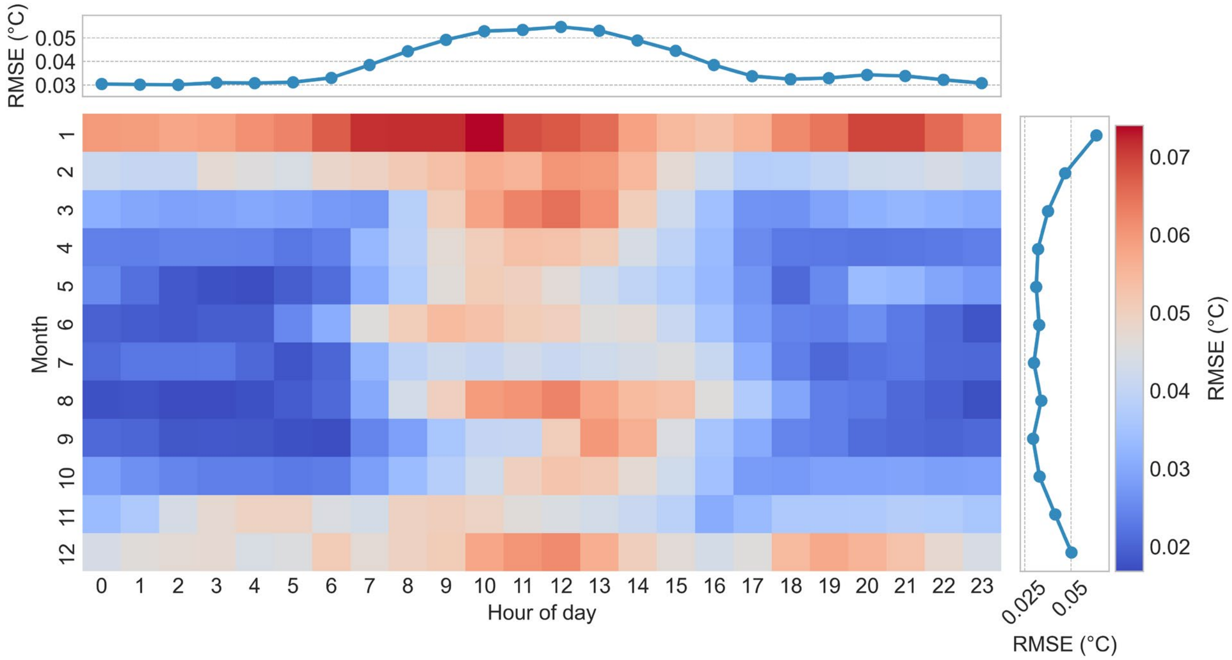

Figure S5. Same as Fig.3, but the RMSEs are normalized by mean values of the corresponding grids.

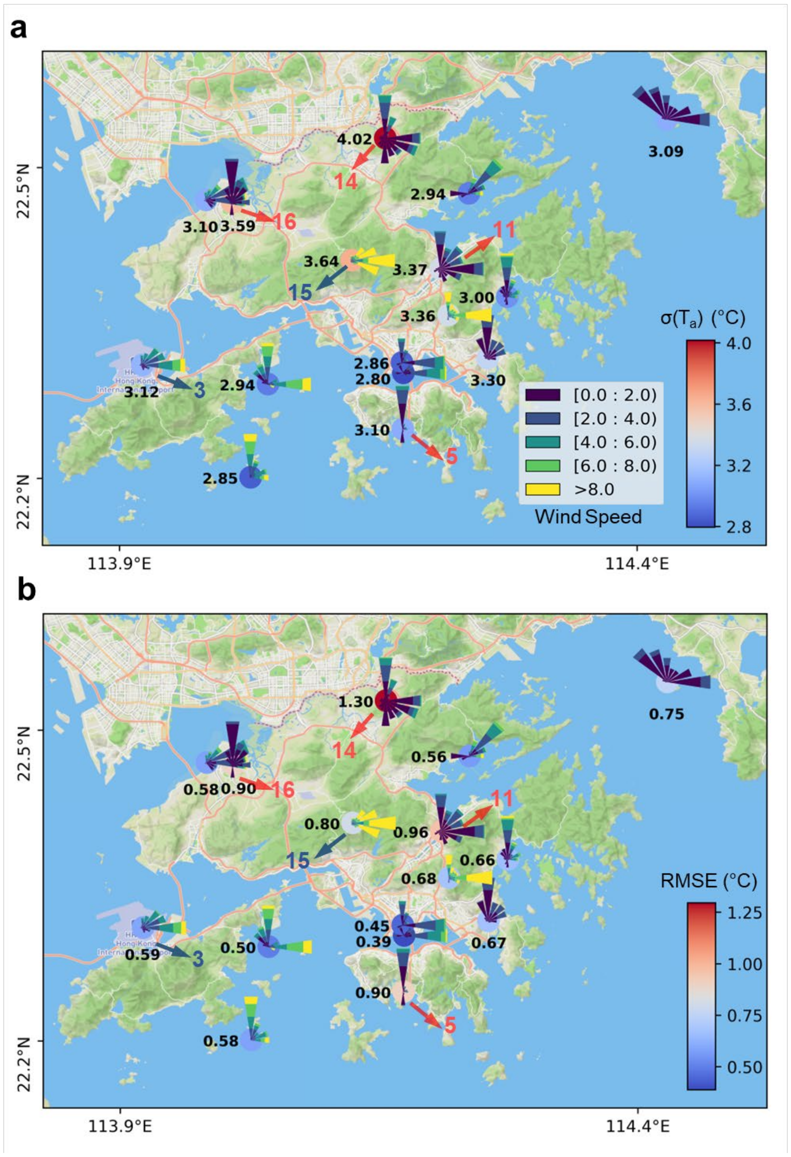

Figure S6. Local wind with corresponding Ta standard deviation **(a)** and forecast errors **(b)** during winter nighttime (basemap © Mapbox). The wind pattern at each site is denoted by Windrose map, where the length denotes the frequency, and color denotes the velocity. Consistent wind distributions in mountains peak (station No.15) and plain airport (station No.3) demonstrate the easterly background wind, while stations No.5, 11, 14 and 16 show larger variances and forecast errors with northern wind.