Li *et al. Geoscience Letters* (2026) 13:11
https://doi.org/10.1186/s40562-026-00460-7

Geoscience Letters

RESEARCH Open Access

A variable-resolution unstructured spatio-temporal graph neural network: an application to very short-range weather forecasting in Guangdong, China

Delin Li[1,2,3], Zhiqiang Chen[1,2,3*], Gang Huang[4,5], Jianjun Xu[2,3] and Yu Zhang[1,2,3]

Abstract
The increase in extreme weather events has raised the demand for very short-range forecasts in Guangdong, the most populous province in China. Although global AI-based weather forecasting models provide strong medium-range guidance, they remain suboptimal for regional very short-range predictions. To address this, we developed a spatio-temporal graph neural network (STGNN) for Guangdong, featuring an unstructured, variable-resolution graph refined over the Pearl River Delta. The core of the network is 1-D dilated causal temporal convolutions and spatial convolutions on combined static and dynamic graphs. A spatial mean constraint and Laplacian-residual regularization were incorporated into the loss function to enhance physical consistency. Trained on historical reanalysis dataset, the model generates hourly forecasts for multiple near-surface atmospheric variables for the next six hours. Verification against an independent test-set and observations, using multiple metrics, demonstrates high predictive skill and strong spatio-temporal coherence for 2 m temperature and sea-level pressure. For 10 m winds and 2 m humidity, model errors grow larger with lead time. Case studies of an extreme heatwave event in Guangzhou and the post-landfall evolution of Super Typhoon Saola effectively capture the evolution characteristics and spatial patterns, while underestimates the peak temperatures, maximum winds and minimum sea-level pressures. These results support the use of variable-resolution STGNN as a practical approach for very short-range regional forecasting in Guangdong, China.

Keywords Spatio-temporal graph neural network, Very short-range forecasting, Unstructured variable-resolution grid, Guangdong China

*Correspondence:
Zhiqiang Chen
czqiang@gdou.edu.cn
[1]College of Ocean and Meteorology, Guangdong Ocean University, Zhanjiang, China
[2]Shenzhen Institute of Guangdong Ocean University, Shenzhen, China
[3]CMA-GDOU Joint Laboratory for Marine Meteorology and South China, Sea Institute of Marine Meteorology, Guangdong Ocean University, Zhanjiang, China
[4]State Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[5]University of Chinese Academy of Sciences, Beijing 100049, China

Introduction

With the intensification of global warming, the frequency of extreme weather events has increased in recent years, underscoring the urgent need for very short-range weather forecasts (up to 12 h) for timely and actionable local warnings (Scher and Messori 2019). Guangdong Province, situated in a subtropical monsoon climate zone on China's southeastern coast, experiences complex, rapidly changing meteorological conditions due to its unique coastal, mountainous, and densely urbanized terrains (Cheng et al. 2021). The Pearl River Delta (PRD), in particular, as one of the country's most densely populated and economically dynamic regions, is vulnerable to heatwaves, typhoons, and heavy rainfall (Yang et al. 2015). This makes reliable very short-range forecasting essential for disaster mitigation and safeguarding agriculture, infrastructure, and socioeconomic stability (Alley et al. 2019).

Traditional numerical models have played a central role in regional short-range weather forecasting (Bauer et al. 2015) and other geoscience applications (e.g. Wang et al. 2025) over the past years. Mesoscale models such as the Weather Research and Forecasting (WRF) model and the China Meteorological Administration mesoscale model (CMA-MESO) have been widely applied in Guangdong Province, contributing to significant advances (Xin et al. 2023; Feng et al. 2021; Huang et al. 2020). However, their reliance on explicit representation of dynamical and physical processes makes high-resolution simulations computationally expensive, limiting their operational use for real-time applications (Schultz et al. 2021; Brotzge et al. 2023). In addition, short-range forecasting requires highly accurate initial conditions, the assimilation of frequent and dense observations remains challenging (Gustafsson et al. 2018). Moreover, numerical models remain inadequate in resolving the fine-scale land-sea-air interactions in Guangdong Province (Jin et al. 2023).

In recent years, the success of deep learning has led to a new generation of data-driven "AI models" that exploit vast historical datasets for "end-to-end" weather prediction (Chen et al. 2023; Chen et al. 2025; Jiménez-Esteve et al. 2025; Ren et al. 2021) and air quality prediction (Fu et al. 2025). Representative examples include GraphCast, developed by Google DeepMind using graph neural networks (GNN) (Lam et al. 2023); FourCastNet, built on Fourier neural operators (FNO) by NVIDIA (Pathak et al. 2022); Pangu-Weather, introduced by Huawei Cloud in 2023, which uses a three-dimensional (3D) Transformer architecture and has achieved medium-range forecast skill comparable to the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS) (Bi et al. 2023; Hua et al. 2025); and, more recently, Microsoft's Aurora AI weather model has shown strong performance in forecasting tropical cyclone, and with fine-tuning, extends its capability to global forecasts of air quality and ocean waves (Bodnar et al. 2025).

In early 2025, ECMWF officially operationalized its AI-based weather model, AIFS, running it alongside the traditional IFS to deliver global medium-range forecasts up to 15 days, marking a milestone in integrating AI into operational weather forecasting (Lang et al. 2024). In July 2025, ECMWF operationally launched the ensemble version of its AI Forecasting System (AIFS-ENS) which consists of 51 perturbed members designed to generate probabilistic distributions. Compared with the physics-based system, AIFS improves forecast accuracy for variables such as surface temperature (Pan et al. 2025). More importantly, benefiting from advanced Graphics Processing Unit (GPU) technology, it produces forecasts more than an order of magnitude faster while reducing energy consumption by roughly three orders of magnitude.

Although AI-based weather models have achieved notable success, they are primarily designed for global medium-range forecasting, typically with a six-hour time step and thus cannot meet the requirements of very short-range (e.g., hourly) predictions (Conti 2024). Moreover, they typically provide outputs on a regular global grid of 0.25° horizontal resolution, which limits their ability to capture fine-scale, unstructured local meteorological features (Camps-Valls et al. 2025). As current progress in regional short-range forecasting remains limited, advancing deep learning–based approaches is therefore essential to address these deficiencies (Xu et al. 2025).

In this study, we propose a variable-resolution unstructured spatio-temporal graph neural network (STGNN) framework that integrates both static and dynamic graphs. The model is trained on historical reanalysis data to capture multivariate, unstructured spatio-temporal patterns and generate six-hourly surface weather forecasts. Its performance is evaluated against independent observational and reanalysis datasets using multiple metrics, and two case studies—including an extreme heat event and a typhoon—are examined to assess the model's forecasting skill.

Data and methods

Spatio-temporal graph neural network (STGNN)

STGNN is a deep learning framework that combines graph structures with time-series modeling via graph convolution for spatial topology and 1-D convolutions for temporal dynamics, to capture spatial and temporal dependencies in non-Euclidean space (Zhou et al. 2020; Sahili and Awad 2023). In this study, we construct an unstructured graph with 250 nodes across Guangdong Province, with increased resolution in the Pearl River Delta (PRD) region (Fig. 1a). This "variable-resolution

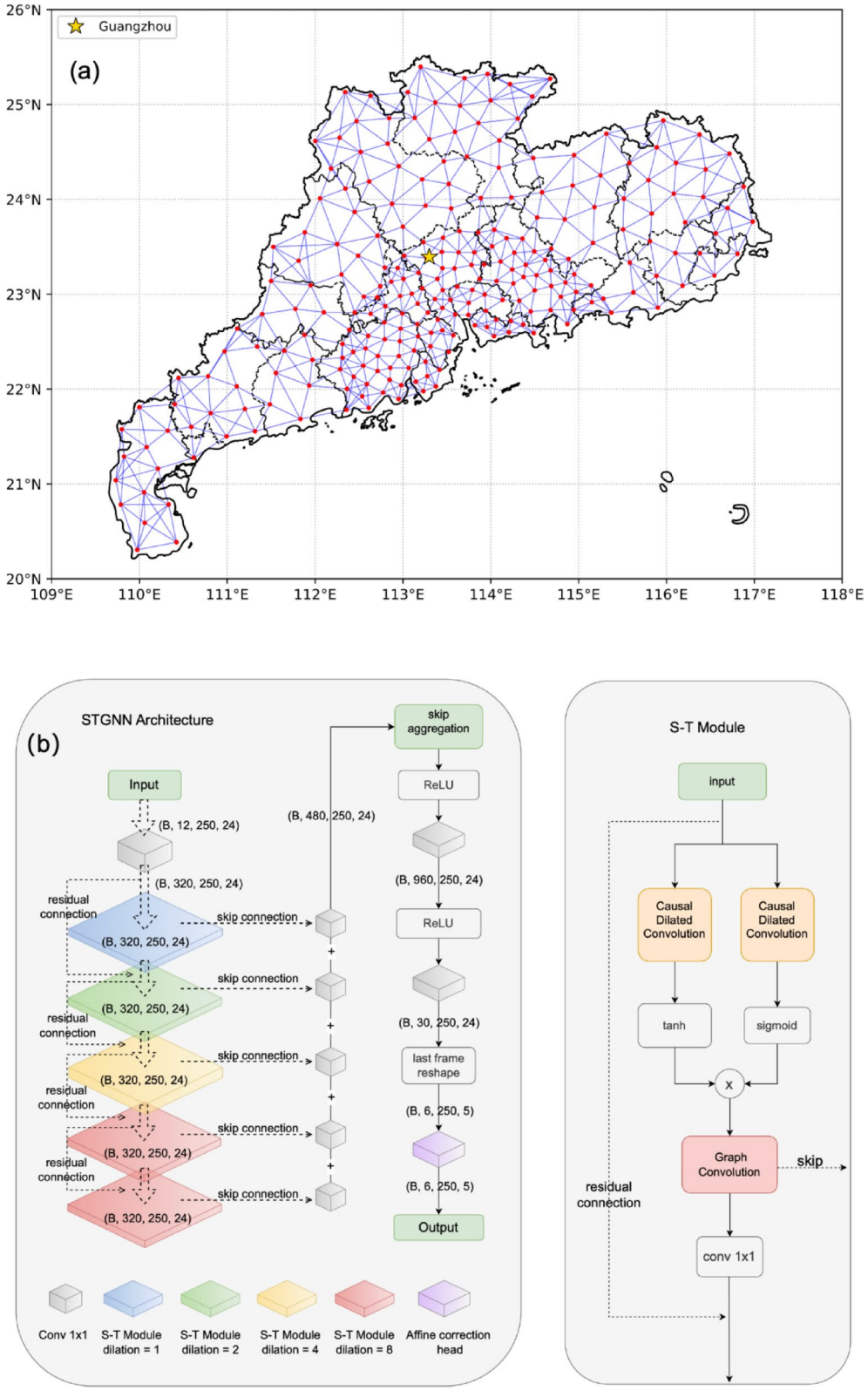

Fig. 1 (**a**) The unstructured graph across Guangdong Province, consisting of 250 nodes with higher resolution in the Pearl River Delta (~ 14 km) to enhance detail in key subregions and coarser resolution elsewhere (~ 28 km) to maintain manageable computational cost. (**b**) The network architecture (left) and the spatio-temporal core module (right)

unstructured graph" simply refers to a spatial mesh in which station density varies by region and nodes are not arranged on a regular latitude–longitude lattice. Our design preserves broad regional coverage while enhancing representation in critical subregions.

The overall architecture builds on Graph WaveNet with some specific modifications (Wu et al. 2019). For temporal modeling, causal dilated convolutions are employed to capture long-range dependencies efficiently. In the spatial dimension, a dynamic graph is introduced alongside the static unstructured graph, enabling graph convolution to represent both fixed topologies and time-varying relationships. The "dynamic graph" denotes adjacency matrices that evolve with time or input features, enabling the network to learn transient, flow-dependent spatial dependencies beyond the static graph. This architecture effectively captures spatio-temporal evolution in multivariate, unstructured meteorological networks, balancing predictive accuracy with computational efficiency and showing strong potential for regional very short-range weather forecasting. A schematic overview of the network architecture is shown in Fig. 1b.

Data

1) **ERA5:** This study uses the fifth-generation global reanalysis dataset (ERA5), produced by the ECMWF. ERA5 combines a wide range of global observations with the IFS model, providing a physically consistent, multivariate dataset coupling the atmosphere, land, and ocean (Hersbach et al. 2020). It resolves 137 vertical model levels from the surface to 0.01 hPa and has a horizontal resolution of approximately $0.25° \times 0.25°$ globally. The dataset is available hourly from January 1940 to the present and is updated in near real time. With its high spatio-temporal consistency and accuracy across a wide range of variables, ERA5 has become one of the most widely used reanalysis products in meteorological and climate research and serves as a standard benchmark for machine learning–based forecasting models (Huang et al. 2024).
2) **ISD-Lite:** We also use the Integrated Surface Database (ISD) from the U.S. National Centers for Environmental Information (NCEI). The ISD aggregates hourly and synoptic observations from meteorological stations worldwide, drawing on digitized historical records as well as modern automated systems. Spanning from 1901 to the present, it provides variables including temperature, dew point, sea-level pressure, wind speed and direction, wind gusts, cloud cover, visibility, precipitation and snow depth (Smith et al. 2011). A simplified version, ISD-Lite, retains commonly used variables while excluding sub-hourly records and redundant information to facilitate processing.
3) **GLO-90:** We use the Copernicus global digital surface elevation model GLO-90 (Copernicus DEM GLO-90), provided under the Copernicus Program by the European Space Agency (ESA) and the European Union (EU). This dataset offers global land coverage (excluding oceans) at a horizontal resolution of approximately 90 m (Guth et al. 2021). Surface elevation values within the study domain are extracted with high fidelity, allowing accurate representation of complex topographic features.

Model training

We use ERA5 data from 2000 to 2022 as the model training set (20% for validation), and employ 2023–2024 as an independent evaluation dataset. Original gridded variables are interpolated to unstructured graph nodes. Model inputs consist of three categories:

1) **Meteorological variables:** hourly fields of five near-surface variables over the past 24 h, including 2 m temperature (T2m), relative humidity (RH, derived from 2 m dew point temperature), mean sea-level pressure (MSLP), and 10 m u- and v-wind components (U10 and V10).
2) **Temporal features:** hour of day and day of year, represented by sine and cosine transformations to preserve periodicity.
3) **Static information:** latitude and longitude, elevation, and the adjacency matrix of the static unstructured graph.

The model predicts the same five variables at hourly intervals for the next six hours. Training samples are generated with a sliding-window approach, and all variables are standardized.

For model training, we employ a composite loss function consisting of three components: Huber loss, spatial mean constraint, and Laplacian residual regularization:

$$L = \mathcal{L}_{\text{Huber}} + \lambda_{\text{space}}\, \mathcal{L}_{\text{space}} + \lambda_{\text{lap}}\, \mathcal{L}_{\text{lap}} \tag{1}$$

The Huber loss $\mathcal{L}_{\text{Huber}}$ is defined as:

$$\mathcal{L}_{\text{Huber}} = \frac{1}{|\Omega|} \sum_{i \in \Omega} \rho_\delta\left(E_i\right) \tag{2}$$

$$\rho_\delta\left(e\right) = \begin{cases} \frac{1}{2} e^2 & |e| \le \delta \\ \delta\,|e| - \frac{1}{2}\delta^2 & |e| > \delta \end{cases} \tag{3}$$

where, $\rho_\delta\left(e\right)$ is the Huber function, $E = \hat{Y} - Y$ is the residual between the prediction and ground truth, Ω

denotes the set of all sample indices, $\delta = 1.0$ is the threshold.

The spatial mean constraint $\mathcal{L}_{\text{space}}$ is designed to penalize discrepancies in spatial averages between prediction and ground truth:

$$\mathcal{L}_{\text{space}} = \text{mean}_N\left(\hat{Y}\right) - \text{mean}_N\left(Y\right)_2^2 \tag{4}$$

where, $\text{mean}_N(\cdot)$ denotes averaging over spatial nodes. The weight is $\lambda_{\text{space}} = 1.0\times\text{e}^{-3}$.

The Laplacian residual regularization encourages residuals to be spatially smooth by penalizing high-frequency variations on the graph, and is defined as:

$$\mathcal{L}_{\text{lap}} = \frac{1}{|\Omega|}\sum_{i\in\Omega}\mathbf{r_i}^{\top}\mathbf{L_r}\mathbf{r_i} \tag{5}$$

where, $\mathbf{r} \in \mathbf{R}^{\boldsymbol{N}}$ denotes the residual vector over all N graph nodes, $\mathbf{L_r}$ is the normalized graph Laplacian constructed from the static adjacency matrix. The weight is $\lambda_{\text{lap}} = 5.0\times\text{e}^{-4}$.

Overall, the Laplacian-residual term suppresses non-physical small-scale noise during training, while the spatial-mean constraint prevents domain-wide drift in the predicted fields. Together, they reduce systematic errors and guide the model toward more physically consistent and stable solutions.

Metrics

To evaluate the model forecasting performance, we use multiple evaluation metrics:

1) **MAE/RMSE:** The Mean Absolute Error (MAE) measures the average absolute error; The Root Mean Squared Error (RMSE) quantify the level of error dispersion, with greater sensitivity to large deviations. Both retain the units of the variable.

$$\begin{aligned} MAE &= \frac{1}{\Omega}\sum_{i=1}^{\Omega}\left|Y_i - \widehat{Y}_i\right| \\ RMSE &= \sqrt{\frac{1}{\Omega}\sum_{i=1}^{\Omega}\left(Y_i - \widehat{Y}_i\right)^2} \end{aligned} \tag{6}$$

2) **$\mathbf{R}^2$Score:** The Coefficient of determination (R^2) is used for evaluating the model six-hourly temporal prediction skill. It quantifies the proportion of variance in the true values explained by the six-hourly forecast relative to a mean-only baseline. Values closer to one indicate better model performance.

$$\text{R}^2 = 1 - \frac{\sum_{i=1}^{\Omega}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{\Omega}\left(Y_i - \hat{Y}\right)^2} \tag{7}$$

3) **SSIM Score:** We use the Structural Similarity Index (SSIM) to evaluate model spatial forecast skill. It measures the similarity of spatial structure between model forecast and truth. SSIM typically ranges from zero to one, with values closer to 1 indicating higher spatial similarity.

$$SSIM = \frac{\left(2\mu_{\hat{Y}}\mu_Y + C_1\right)\left(2\sigma_{\hat{Y}Y} + C_2\right)}{\left(\mu_{\hat{Y}}^2 + \mu_Y^2 + C_1\right)\left(\sigma_{\hat{Y}}^2 + \sigma_Y^2 + C_2\right)} \tag{8}$$

where, $\mu_{\hat{Y}}$ and μ_Y denote the mean values of the forecast and truth, $\sigma_{\hat{Y}}^2$ and σ_Y^2 are their variances. $\sigma_{\hat{Y}Y}$ represents the covariance and C_1 and C_2 are small constants introduced to ensure numerical stability.

Results

Spatial evaluation of model forecast skill

We first examined the spatial distribution of the model forecast performance across the whole evaluation dataset using MAE, RMSE and R^2 (Fig. 2). The test set is further divided into four seasonal subsets, and MAE/RMSE are computed separately for each subset, as summarized in Table 1. The first column presents the mean fields of five near-surface meteorological variables from ERA5. The second and third columns show the mean MAE and RMSE of the model forecasts. Overall, MAE and RMSE exhibit consistent spatial patterns across different variables, suggesting that the model forecasts are stable and minimally affected by outliers. For 2 m temperature and humidity, errors are concentrated in the eastern region, whereas the Pearl River Delta shows the smallest errors, with mean values of approximately 0.75 °C and 4.4%, respectively. Errors in mean sea-level pressure remain close to 0.6 hPa, with slightly larger values in the north. In contrast, the 10 m horizontal wind components (u and v) show greater errors along the southern coast (with maximum over 1 m/s), where wind structures are mainly driven by complex land-sea interactions, while errors in the northern area are comparatively smaller (about 0.5 m/s). We further assessed forecast skill over a 6-h horizon using the coefficient of determination (R^2). For 2 m temperature and mean sea-level pressure, R^2 values are generally above 0.7 (locally exceeding 0.8), indicating relatively high predictive skill. Humidity yields slightly lower skill, though R^2 remains above 0.6 overall. Wind forecasts show the weakest performance, with R^2 values around 0.4, consistent with their higher error levels.

Temporal evaluation of model forecast skill

We further examined how model forecast performance evolves with lead time. All variables were standardized

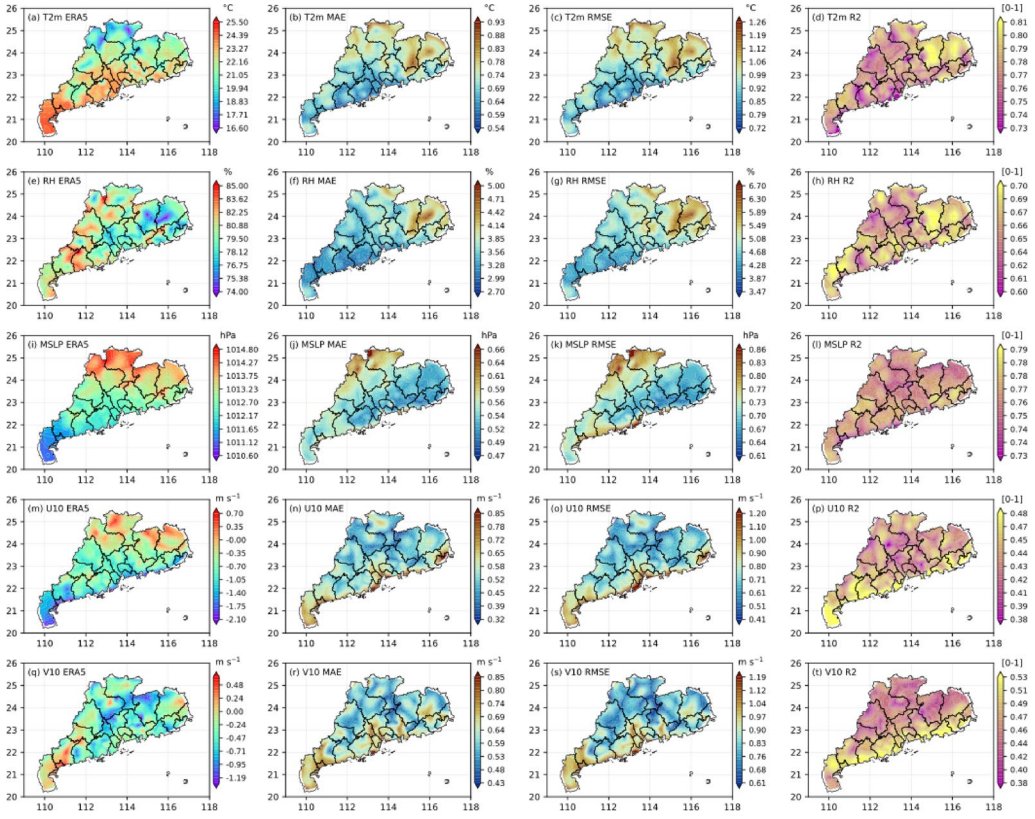

Fig. 2 Spatial distribution of the mean fields from ERA5 (first column), along with model forecast errors evaluated by MAE (second column), RMSE (third column), and R^2 (fourth column). Panels (**a**–d) show results for 2 m temperature, (**e**–h) for 2 m relative humidity, (**i**–**l**) for mean sea-level pressure, (**m**–**p**) for the 10 m u-wind component, and (**q**–**t**) for the 10 m v-wind component

Table 1 MAE / RMSE of the test set for different seasonal periods

	Jan-Feb-Mar	Apr-May-Jun	Jul-Aug-Sep	Oct-Nov-Dec
T2m (℃)	0.73/0.96	0.60/0.82	0.58/0.79	0.56/0.74
RH (%)	4.27/5.55	2.86/3.94	3.11/4.18	3.21/4.25
MSLP (hPa)	0.44/0.56	0.59/0.76	0.45/0.57	0.37/0.49
U10 (m/s)	0.48/0.65	0.54/0.74	0.49/0.64	0.46/0.62
V10 (m/s)	0.53/0.72	0.64/0.88	0.54/0.70	0.50/0.68

to enable inter-comparison. The standardized MAE and RMSE were calculated as a function of forecast horizon. Overall, MAE and RMSE show consistent temporal patterns, suggesting that forecast errors remain stable over time (Fig. 3a and 3b). As expected, errors for all variables increase with lead time, though their magnitudes and growth rates differ. The linear RMSE growth rate for each variable is computed: Mean sea-level pressure (1.37%) and 2 m temperature (2.03%) exhibit the smallest errors and the slowest error growth, reflecting the model's high skill for relatively stable fields. Relative humidity errors fall in the mid-range (3.71%), while 10 m horizontal wind, particularly the u-component (5.78%, 4.39% for v-wind), shows the largest errors, consistent with the spatial error distributions. We also assessed spatial similarity over forecast time using SSIM scores, adopting 0.8 as the threshold for skillful forecasts (Fig. 3c). Within the six-hour lead time, the model effectively reproduces the spatial structures of 2 m temperature and mean sea-level pressure fields. In contrast, for humidity and wind fields, the model maintains reasonable skill during the first three hours, but its ability to capture spatial structures declines thereafter as errors increase.

Case study I: an extreme heat event in Guangzhou

Beyond the evaluation of overall model performance, we carried out case studies on extreme very short-range weather events. From 3 to 8 August 2024, Guangzhou experienced a five-day extreme summer heatwave, with daily maximum 2 m temperatures well above the long-term average for the same period. This event had notable impacts on urban residents, public health, and electricity demand. Station observations from ISD-Lite show that maximum temperatures exceeded 37 °C for three consecutive days, peaking at 38 °C (Fig. 4a). Concurrently, as temperature increased, saturation vapor pressure rose sharply, and minimum relative humidity dropped to around 50% (Fig. 4b). ERA5 data reproduce the general evolution reasonably well but underestimate the maximum temperature (~ 1.5 °C) and overestimate the lowest relative humidity (~ 8%). These biases may be attributed to systematic model errors and uncertainties introduced

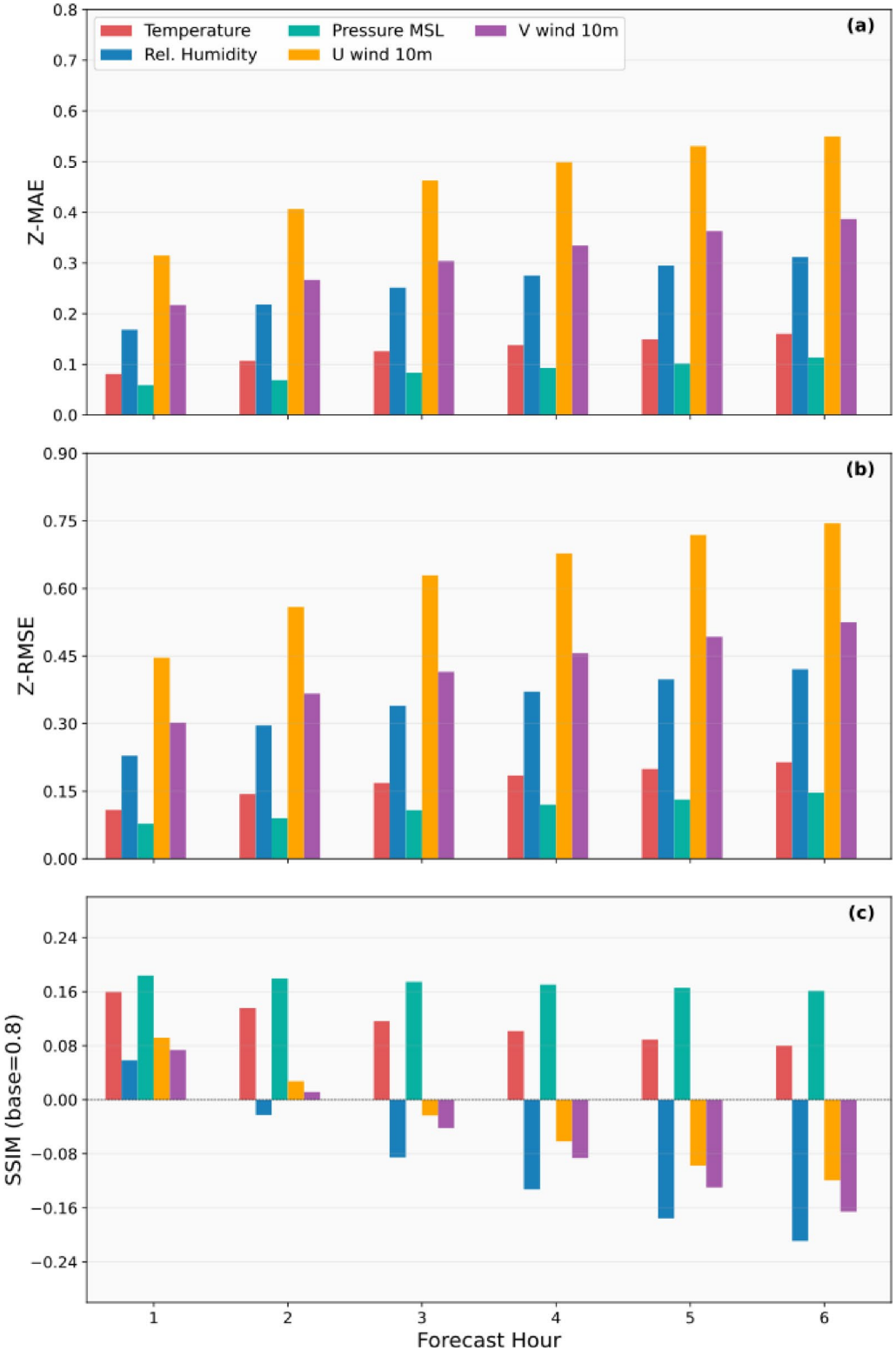

Fig. 3 Temporal evaluation of forecast skill for five near-surface variables. (**a**) Standardized MAE (Z-Score). (**b**) Standardized RMSE (Z-Score). (**c**) SSIM relative to a skill threshold of 0.8

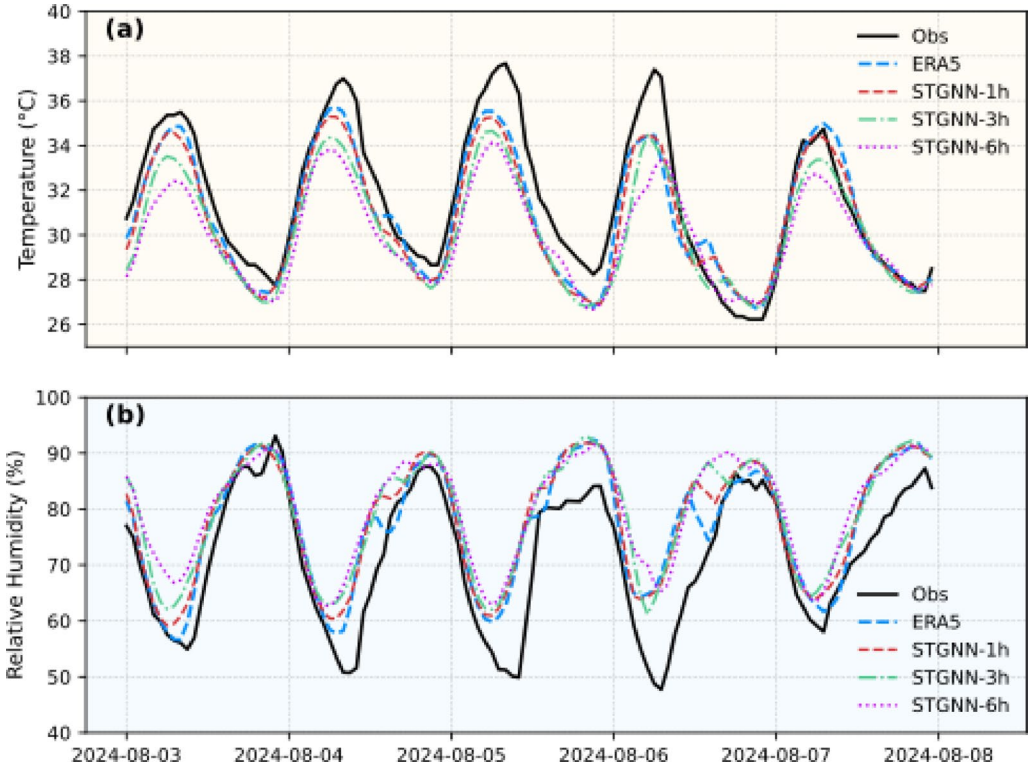

Fig. 4 Time series of observed (Obs), ERA5, and model forecasts (1 h, 3 h, and 6 h lead) for (**a**) 2 m temperature and (**b**) relative humidity in Guangzhou during the extreme heatwave from 3 to 8 August 2024

from station interpolation. The one-hour lead model forecasts agree closely with ERA5, thus inheriting these biases relative to the observations. At three- and six-hour lead times, these biases become larger, especially for extreme values (Fig. 4). Even so, the model overall gives a reasonable representation of the heatwave event.

Case study II: Super Typhoon Saola

Guangdong Province is among the regions in China most frequently affected by tropical cyclones, which regularly trigger very short-range extreme weather and cause severe societal and economic impacts. To evaluate the model's capability for such events, we selected Super Typhoon Saola in 2023 as a representative case. Saola was one of the strongest typhoons to make landfall in Guangdong in recent years. According to reports by the Hong Kong Observatory, Saola originated over the western Pacific on 24 August, strengthened as it moved northwestward into the South China Sea, and reached typhoon intensity before landfall on the coast of Zhuhai at round 20:30 (UTC) on 1 September. It produced destructive winds and heavy rainfall across the Pearl River Delta.

We examined the six-hour post-landfall evolution of Typhoon Saola. ERA5 reanalysis data show that the cyclone continued moving westward, maintaining a minimum central pressure of around 980 hPa and maximum sustained winds exceeding 20 m/s (Fig. 5a). The model forecasts reproduced these features with reasonable accuracy (Fig. 5b). During the first three hours, the central pressure was slightly underestimated and wind speed errors remained small. Beyond three hours, pressure and wind biases became more pronounced. This suggests that for strong typhoons, the model's forecasting skill remains limited but still shows a certain level of predictive capability at short lead times.

Discussion

The above results indicate that the model provides a reasonable level of forecast skill. To further assess its performance, we adopted a 6-h persistence forecast as a baseline and compared the RMSE of the STGNN with that of the Pangu-Weather forecasts (Fig. 6). The STGNN consistently outperforms the Pangu model across all lead times. At the 1-h lead time, its performance is comparable to the persistence baseline; however, as lead time increases to 2–6 h, the persistence forecast degrades rapidly, whereas the STGNN exhibits only a modest increase in error. These findings highlight an important limitation: although state-of-the-art AI weather models excel at global medium-range prediction, their skill does not necessarily transfer to localized, high-resolution very short-term forecasting. This underscores the need to develop dedicated AI models tailored for regional very short-term prediction.

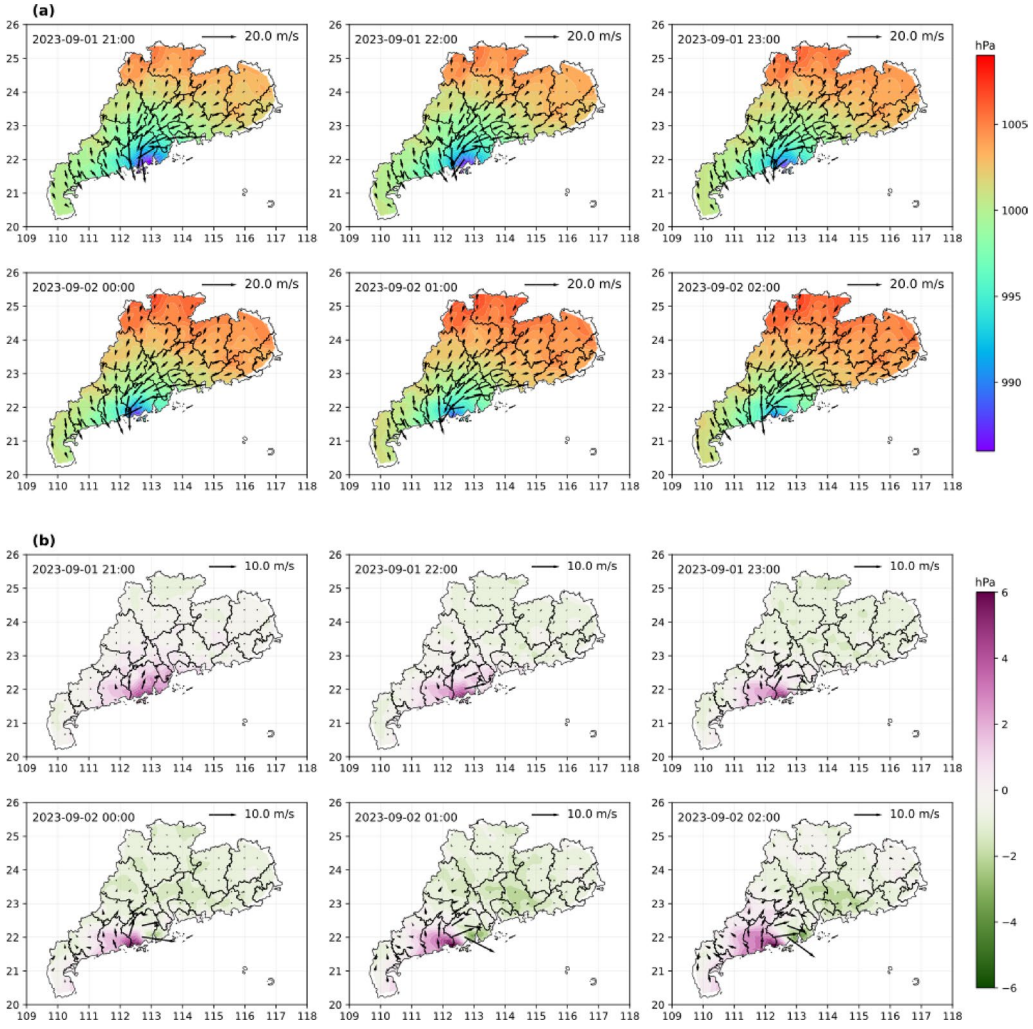

Fig. 5 Time evolution of Typhoon Saola from 21:00 on 1 September to 02:00 on 2 September 2023. (**a**) ERA5 reanalysis of mean sea-level pressure (shaded, hPa) with 10 m wind vectors. (**b**) Model forecast biases relative to ERA5 for sea-level pressure (shaded, hPa) and 10 m winds (vectors)

Forecasts of winds show larger errors that grow with forecast lead time. These degradations are especially pronounced in the southern coastal regions, where strong interactions among large-scale circulation, complex topography, and heterogeneous surface conditions. These factors produce rapidly varying wind structures that are difficult for the model to capture. The temporal evolution of spatial similarity (Fig. 3c) also shows discrepancies in relative humidity, which is strongly influenced by surface conditions, evapotranspiration, and wind variability. As a result, errors in wind forecasts propagate into the humidity field, amplifying structural differences. Given the heterogeneous land-surface and topographic features, incorporating additional static (e.g., land-use, land-type) or dynamic (e.g., soil moisture, vegetation state) information into the predictors could enhance the model's ability and performance. In addition, the STGNN is a general framework; with appropriate retraining or fine-tuning that accounts for regional topography and surface conditions, the model can be transferred to other areas.

Although the model succeeds in reproducing the overall evolution of extreme events, its ability to forecast the peaks (or troughs) remains limited (Figs. 4 and 5). The ERA5 reanalysis data, which are produced with the ECMWF IFS global numerical model, are well-suited for mid-range forecasting but are likely to underrepresent small-scale extremes (Potisomporn et al. 2023). Previous studies have also shown that current numerical models tend to underestimate short-range extreme weather events, such as intense tropical cyclones, heatwaves, and heavy rainfall (Ghimire et al. 2019; Hsu et al. 2019; Done et al. 2015). Incorporating observational datasets into the training process, or applying observationally constrained fine-tuning to the pre-trained AI models, may provide a promising pathway to further improve forecasting skill.

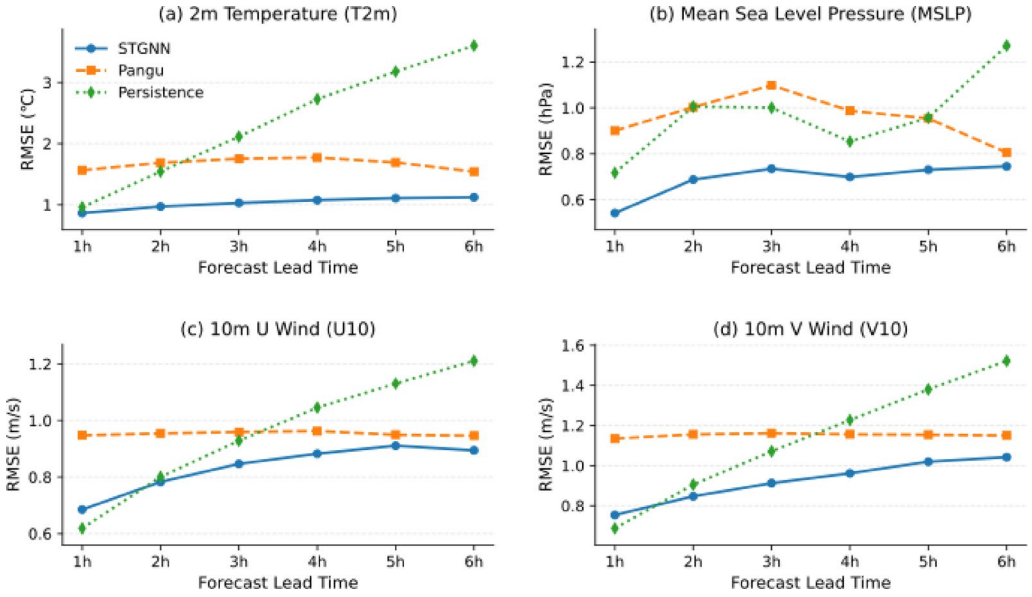

Fig. 6 Six-hour RMSE evolution for the STGNN, Pangu-Weather, and persistence models: (**a**) 2 m temperature; (**b**) mean sea-level pressure; (**c**) 10 m u-wind; (**d**) 10 m v-wind

Conclusion

This study developed a variable-resolution unstructured spatio-temporal graph neural network (STGNN) for Guangdong Province, with refined grids over the Pearl River Delta. The model combines static and dynamic graphs to capture the coordinated spatial evolution of multiple meteorological variables. Spatial consistency was reinforced by incorporating a spatial mean constraint and Laplacian residual regularization into the loss function. Trained on multi-year historical ERA5 reanalysis, the model generates six-hourly forecasts of 2 m temperature, mean sea-level pressure, relative humidity, and 10 m wind. Evaluation against independent observations and ERA5 test sets showed relatively low errors and high spatial–temporal consistency for temperature and pressure, with R^2 and SSIM values reaching around 0.8. Forecast skill degraded with lead time for wind and humidity, with the wind R^2 score dropping to around 0.5 and humidity SSIM score decreasing to approximately 0.6 at six-hour lead time. These results highlight the challenges of capturing rapidly evolving wind structures and the associated wind–evaporation feedback on humidity. Additional case studies with extreme events, including persistent heatwaves and a landfalling strong typhoon, further revealed that the model can reproduce the overall evolution of such extremes but tends to underestimate their intensity, particularly the peak temperatures during heatwave and the strong winds associated with severe tropical cyclones.

In summary, this work demonstrates the potential of spatio-temporal graph-based deep learning for very short-range regional weather forecasting, while also identifying critical challenges related to surface heterogeneity, wind–humidity coupling, and extreme-event intensity. Incorporating detailed surface parameters, and applying observationally constrained fine-tuning, could represent promising directions for improving model performance.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (42206253, 42205025, 42130605, 42141019, 42261144687, 42575170), the program for scientific research start-up funds of Guangdong Ocean University (060302032305, 060302032105), and the Hainan Key Research and Development Project under Grant (ZDYF2023SHFZ125).

Author contributions

D. L., Z. C., G. H., and Y. Z. conceived the research. Z. C. and D. L. performed the research. D. L., Z. C., G. H., and J. X. wrote the paper. All authors reviewed the manuscript.

Data availability

ERA5 reanalysis (single-level) data are provided by the Copernicus Climate Data Store: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels? tab=download. Surface hourly meteorological observations are available from the NOAA National Centers for Environmental Information: https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso? id=gov.noaa.ncdc: C00532. GLO-90 elevation data are available from OpenTopography: https://portal.opentopography.org/raster? opentopoID=OTSDEM.032021.4326.1, and can also be accessed via the Open-Meteo Elevation API: https://open-meteo.com/en/docs/elevation-api. The source code and the model output data can be accessed at https://doi.org/10.5281/zenodo.17074830. Pangu-Weather data can be downloaded from https://aws.amazon.com/marketplace/pp/prodview-cjmveqlu2ttkk#resources.

Declarations

Competing interests

The authors declare no competing interests.

Received: 19 September 2025 / Accepted: 14 January 2026

Published online: 09 February 2026

References

Alley RB, Emanuel KA, Zhang F (2019) Advances in weather prediction. Science 363:342–344. https://doi.org/10.1126/science.aav7274

Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. Nature 525:47–55. https://doi.org/10.1038/nature14956

Bi K, Xie L, Zhang H et al (2023) Accurate medium-range global weather forecasting with 3D neural networks. Nature 619:533–538. https://doi.org/10.1038/s41586-023-06185-3

Bodnar C, Bruinsma WP, Lucic A et al (2025) A foundation model for the Earth system. Nature 641:1180–1187. https://doi.org/10.1038/s41586-025-09005-y

Brotzge JA, Berchoff D, Carlis DL et al (2023) Challenges and opportunities in numerical weather prediction. Bull Am Meteorol Soc 104:E698–E705. https://doi.org/10.1175/BAMS-D-22-0172.1

Camps-Valls G, Fernández-Torres M-Á, Cohrs K-H et al (2025) Artificial intelligence for modeling and understanding extreme weather and climate events. Nat Commun 16:1919. https://doi.org/10.1038/s41467-025-56573-8

Chen L, Zhong X, Zhang F et al (2023) FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. NPJ Clim Atmos Sci 6:190. https://doi.org/10.1038/s41612-023-00512-1

Chen K, Han T, Ling F et al (2025) The operational medium-range deterministic weather forecasting can be extended beyond a 10-day lead time. Commun Earth Environ 6:518. https://doi.org/10.1038/s43247-025-02502-y

Cheng J, Bambrick H, Yakob L et al (2021) Extreme weather conditions and dengue outbreak in Guangdong, China: spatial heterogeneity based on climate variability. Environ Res 196:110900. https://doi.org/10.1016/j.envres.2021.110900

Conti S (2024) Artificial intelligence for weather forecasting. Nat Rev Electr Eng 1:8–8. https://doi.org/10.1038/s44287-023-00009-2

Done JM, Holland GJ, Bruyère CL et al (2015) Modeling high-impact weather and climate: lessons from a tropical cyclone perspective. Clim Change 129:381–395. https://doi.org/10.1007/s10584-013-0954-6

Feng JL, Gao Y, Xia X et al (2021) Impacts of a large-scale adaptive blending scheme for CMA-MESO on regional forecasts a case study of Typhoon Haima. J Trop Meteorol 27:330–345. https://doi.org/10.46267/j.1006-8775.2021.029

Fu Z, Yang X, Ma Y et al (2025) Integrating explainable AI and causal inference to unveil regional air quality drivers in China. J Environ Manage 390:126270. https://doi.org/10.1016/j.jenvman.2025.126270

Ghimire U, Srinivasan G, Agarwal A (2019) Assessment of rainfall bias correction techniques for improved hydrological simulation. Int J Climatol 39:2386–2399. https://doi.org/10.1002/joc.5959

Gustafsson N, Janjić T, Schraff C et al (2018) Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. Q J R Meteorol Soc 144:1218–1256. https://doi.org/10.1002/qj.3179

Guth PL, Van Niekerk A, Grohmann CH et al (2021) Digital elevation models: terminology and definitions. Remote Sens 13:3581. https://doi.org/10.3390/rs13183581

Hersbach H, Bell B, Berrisford P et al (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146:1999–2049. https://doi.org/10.1002/qj.3803

Hsu W-C, Patricola CM, Chang P (2019) The impact of climate model sea surface temperature biases on tropical cyclone simulations. Clim Dyn 53:173–192. https://doi.org/10.1007/s00382-018-4577-5

Hua Z, Hakim G, Anderson-Frey A (2025) Performance of the Pangu-Weather deep learning model in forecasting tornadic environments. Geophys Res Lett 52:e2024GL109611. https://doi.org/10.1029/2024GL109611

Huang Y, Wang Y, Xue L et al (2020) Comparison of three microphysics parameterization schemes in the WRF model for an extreme rainfall event in the coastal metropolitan city of Guangzhou, China. Atmos Res 240:104939. https://doi.org/10.1016/j.atmosres.2020.104939

Huang G, Wang Y, Ham Y-G et al (2024) Toward a learnable climate model in the artificial intelligence era. Adv Atmos Sci 41:1281–1288. https://doi.org/10.1007/s00376-024-3305-9

Jiménez-Esteve B, Barriopedro D, Johnson JE, García-Herrera R (2025) AI-driven weather forecasts to accelerate climate change attribution of heatwaves. Earths Future 13:e2025EF006453. https://doi.org/10.1029/2025EF006453

Jin J, Dong X, He J et al (2023) A regional air-sea coupled model developed for the East Asia and Western North Pacific monsoon region. J Geophys Res Atmos 128:e2022JD037957. https://doi.org/10.1029/2022JD037957

Lam R, Sanchez-Gonzalez A, Willson M et al (2023) Learning skillful medium-range global weather forecasting. Science 382:1416–1421. https://doi.org/10.1126/science.adi2336

Lang S, Alexe M, Chantry M, et al (2024) AIFS -- ECMWF's data-driven forecasting system. https://arxiv.org/abs/2406.01465

Pan L, Zhang H, Liang M et al (2025) Comparative analysis of precipitation forecasts between the ECMWF artificial intelligence system (AIFS) and its integrated forecast system (IFS). Weather Forecast 40:1653–1670

Pathak J, Subramanian S, Harrington P, et al (2022) FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators. https://arxiv.org/abs/2202.11214

Potisomporn P, Adcock TAA, Vogel CR (2023) Evaluating ERA5 reanalysis predictions of low wind speed events around the UK. Energy Rep 10:4781–4790. https://doi.org/10.1016/j.egyr.2023.11.035

Ren X, Li X, Ren K et al (2021) Deep learning-based weather prediction: a survey. Big Data Res 23:100178. https://doi.org/10.1016/j.bdr.2020.100178

Sahili ZA, Awad M (2023) Spatio-Temporal Graph Neural Networks: A Survey. https://arxiv.org/abs/2301.10569

Scher S, Messori G (2019) How global warming changes the difficulty of synoptic weather forecasting. Geophys Res Lett 46:2931–2939. https://doi.org/10.1029/2018GL081856

Schultz MG, Betancourt C, Gong B et al (2021) Can deep learning beat numerical weather prediction? Philos Trans Royal Soc a: Math, Phys Eng Sci 379:20200097. https://doi.org/10.1098/rsta.2020.0097

Smith A, Lott N, Vose R (2011) The integrated surface database: recent developments and partnerships. Bull Am Meteorol Soc 92:704–708. https://doi.org/10.1175/2011BAMS3015.1

Wang T, Fu Z, Zhang S, Li Z (2025) Water erosion risk assessment and predictive modelling for cultural heritage under climate change: a case study of the Great Wall in the Yellow River Basin, China. J Clean Prod 510:145645. https://doi.org/10.1016/j.jclepro.2025.145645

Wu Z, Pan S, Long G, et al (2019) Graph WaveNet for Deep Spatial-Temporal Graph Modeling. https://arxiv.org/abs/1906.00121

Xin R, Li X-X, Shi Y et al (2023) Study of urban thermal environment and local circulations of Guangdong-Hong Kong-Macao Greater Bay Area using WRF and local climate zones. J Geophys Res Atmos 128:e2022JD038210. https://doi.org/10.1029/2022JD038210

Xu P, Zheng X, Gao T et al (2025) An artificial intelligence-based limited area model for forecasting of surface meteorological variables. Commun Earth Environ 6:372

Yang L, Scheffran J, Qin H, You Q (2015) Climate-related flood risks and urban responses in the Pearl River Delta, China. Reg Environ Change 15:379–391. https://doi.org/10.1007/s10113-014-0651-7

Zhou J, Cui G, Hu S et al (2020) Graph neural networks: a review of methods and applications. AI Open 1:57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

Publisher's note