Data-Driven Regional Weather Forecasting Guided by Global Context

NingZi Hu[1*], Guang Yu[2*], Yi Han[2], YingWang[1], XueRong Cui[1], JunXing Zhu[2]✉, and Xiang Wang[2] ✉

[1] China University of Petroleum(East)
{z22160065, z23160106, cuixuerong}@s.upc.edu.cn
[2] National University of Defense Technology
{guangyu, hanyi12, zhujunxing, xiangwangcn}@nudt.edu.cn

Abstract. Weather forecasting plays a critical role in many aspects of modern society. Compared to global weather forecasting, regional weather forecasting (RWF) can achieve higher resolution and more detailed analyses for the target region. Since global weather forms an interconnected system, incorporating the global weather state is crucial for accurate and reliable RWF. However, existing RWF methods suffer from two issues: **(1)** They tend to overlook the influence of global weather patterns on regional weather by limiting the modeling scope to a confined geographic area. **(2)** They usually require extra regional boundary processing to impose lateral boundary forcing for yielding more reasonable and accurate regional forecasts. To address above issues, we propose a global context guided data-driven RWF method that can realize highly effective RWF. First, to our best knowledge, we propose to introduce entire global weather state into RWF modeling for the first time, so as to fully consider the influence of external weather. Second, we design a dual-branch architecture that directly accepts both regional and global weather data as inputs for learning, which eliminates the additional processing on regional boundaries. Finally, we develop a multi-stage regional-global attention mechanism, which fuses regional and global weather features to yield global context, and use it to guide RWF. Experimental results show that the proposed method can achieve state-of-the-art performance and demonstrate the significant improvement in regional forecasting by considering global weather state.

1 Introduction

Weather forecasting is crucial for human safety and societal functioning. While global weather forecasting (GWF) provides a comprehensive overview of atmospheric conditions on a large scale, regional weather forecasting (RWF) delivers more detailed analyses by focusing on smaller geographic area with higher spatial resolution. Consequently, RWF is more effective in predicting localized weather events and addressing the specific forecast requirements of particular region [16].

---

* equal contribution.

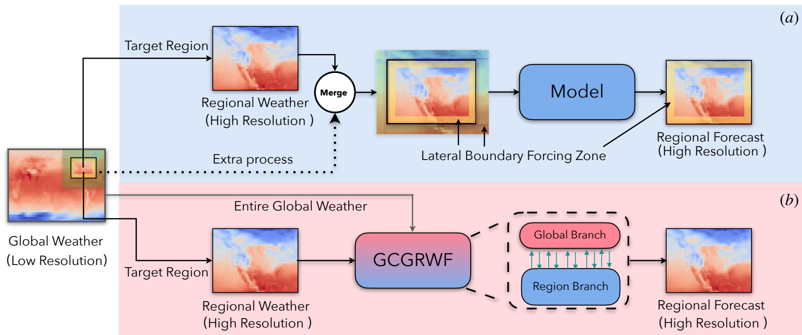

Fig. 1. Comparison between **(a)** existing RWF methods and **(b)** the proposed GCGRWF. Exsiting RWF methods apply lateral boundary forcing to consider limited weather information outside the target region. By contrast, the proposed GCGRWF introduces the entire global weather state into RWF modeling without extra processing to fully consider the influence of external weather state.

It is worth noting that since atmosphere is interconnected [17], weather state outside the target region can significantly affect the weather of the region. Large-scale atmospheric patterns, which originate far beyond the target region, often exert a profound influence on the regional weather. For example, El Niño-Southern Oscillation (ENSO), a weather event originates in the tropical Pacific has far-reaching effects on regional weather patterns of distant areas like North America [18]. Existing RWF methods consider external weather state by lateral boundary forcing, which aims to transfer the weather information outside the region (especially near the regional boundary) to the inside [10, 16].

Since the 1970s, traditional RWF typically adopts Numerical Weather Prediction (NWP) methods. NWP methods treat weather forecasting as an initial value problem [11], which uses the current atmospheric conditions as initial conditions, and integrates a set of partial differential equations (PDEs) to predict future weather state. To realize lateral boundary forcing and consider the external weather state outside the target region, regional NWP methods adopt lateral boundary conditions(LBCs), which consist of a set of well-posed equations for introducing external weather information [8]. These equations are typically calculated using data around the regional boundary, i.e. the data points in the green part of lateral boundary forcing zone shown in Fig. 1 (a).

With the rapid development of deep learning, RWF has utilized deep neural networks (DNNs) for modeling and prediction, and achieved promising performance [1, 6]. In contrast to regional NWP methods, which depend on solving PDEs and involve substantial computational overhead, DNN based RWF methods identify patterns and regularities within weather data to achieve forecasting. Compared with NWP methods, DNN based methods can achieve comparable forecasting performance at a much faster inference speed [1]. Similarly, DNN based RWF methods follow the practice of regional NWP methods to apply lateral boundary forcing. Instead of using formula-based LBCs for lateral boundary forcing, DNN based RWF methods usually impose extra processing on the regional boundaries (like smoothing or replacing the boundary forecasts on yellow part in Fig. 1 (a), detailed in Sec.2), so as to make the weather forecast of the region more accurate and more in line with the external weather patterns.

Although RWF has made considerable progress, both regional NWP methods and DNN based RWF methods still suffer from two issues: **(1)** They only take a limited scope of external weather state into account, and often neglect broader global-scale weather patterns. As discussed previously, existing RWF methods introduce the external weather state through lateral boundary forcing. However, this way still limits the modeling scope to a constrained spatial area, and it can only introduce the external weather information around the regional boundaries. **(2)** They usually require extra regional boundary processing to apply lateral boundary forcing. Due to the different spatial resolutions of global weather data and regional weather data, regional NWP methods need to manually align their resolutions (such as interpolating coarse-resolution global data) before implementing LBCs [8, 15]. As for DNN based RWF methods, they often require additional processing on regional boundaries to ensure the reliability of regional or boundary weather forecasts.

To address above issues, we propose a global context guided data-driven RWF (GCGRWF) method, which can achieve more effective and accurate RWF by integrating global weather into RWF. To our best knowledge, this is the first work to consider entire global weather in RWF. Specifically, this paper contributes to RWF in terms of the following three aspects:

- Unlike existing RWF methods that limit their modeling area within a constrained spatial range, we propose to introduce the global-scale weather state into the regional forecasting modeling, thereby fully considering the influence of external weather patterns on the target region for effective RWF.
- Instead of extra processing on regional boundary to introduce external weather information, we propose a dual-branch DNN based architecture, in which each branch takes regional and entire global weather data as input respectively for learning, so as to seamlessly integrate them.
- To achieve effective global information guidance, we design a multi-stage regional-global feature attention mechanism. It first fuses global features and regional features to yield global context, which is then embedded into the attention mechanism to capture the global information.

Experimental results demonstrate that the proposed method is able to outperform all compared RWF methods by a substantial margin and highlight that incorporating global weather context significantly enhances RWF accuracy.

2 Related Works

DNN based RWF methods learn to predict future regional weather by taking current weather as input of DNN. Once the DNN is trained, forecast results can be obtained quickly through one forward pass of DNN. For example, the inference cost of Pangu-Weather takes merely 1,400ms on a single GPU, which is 10000x faster than operational Integrated forecasting system (IFS) [1]. To realize effective RWF, various DNN architectures have been explored, such as convolution neural network (CNN), graph neural network (GNN), and transformer. Following the practice of regional NWP methods, lateral boundary forcing is also applied in DNN based RWF methods. Specifically, Oskarsson et al. are inspired by a GNN based GWF model (i.e. GraphCast [6]), and propose a RWF method named Hi-LAM [10]. Hi-LAM utilizes a hierarchical mesh graph to mitigate circular artifacts in prediction outputs. To apply boundary forcing and improve the accuracy of forecasts, Hi-LAM directly replaces the prediction boundary with ground truth data (i.e. reanalysis data), and then used as input for subsequent training iterations. Motivated by Mamba, a state-of-the-art state-space model that is good at capture long-term dependencies, Qin et al. propose MetMamba [12] for RWF. MetMamba deploys two training stages. It first uses regional weather data to pre-train a RWF DNN. To incorporate lateral boundary forcing, it then employs LBC adaptive auto-regressive training, which aims to replace data points on regional prediction boundary with prediction from a GWF model in training. Ying Long [15] utilizes window attention from the Swin Transformer [7] for local feature extraction and the Adaptive Fourier Neural Operator(AFNO) module [5] to capture global features in target region. Lateral boundary forcing is not applied during the training phase. However, it smooths the boundaries of its predictions by blending them with NWP forecasts. Specifically, the boundaries of the forecast are replaced with the average of the DNN based method's forecast and the NWP method's forecast.

As discussed above, the current RWF methods, including regional NWP methods and DNN based RWF methods, need to process regional boundaries additionally to introduce external weather information for more reasonable and accurate RWF. Meanwhile, none of them fully consider the global weather state that may have significant impacts on regional weather.

3 The Proposed Method

As shown in Fig. 2, we propose a global context guided RWF method named GCGRWF, which smoothly integrates entire global weather into the DNN based RWF modeling. The details of GCGRWF are depicted as follows:

3.1 Light-Heavy Dual Branches

To avoid extra operations on regional boundary and facilitate the introduction of external weather information, we design a dual-branch DNN based architecture to separately process coarse-resolution global weather data and high-resolution regional data at the model input stage. With the powerful automatic feature learning ability of DNN, we no longer need to manually process global and regional weather data with different resolutions before forecasting. Meanwhile,

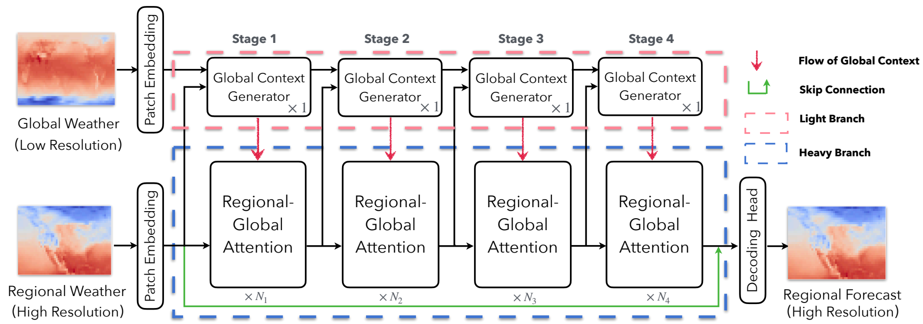

Fig. 2. Overall framework of the proposed method (GCGRWF).

since we fully consider global weather to improve the reliability and rationality of RWF, additional operations on regional boundary (e.g. smoothing) in DNN based RWF methods are no longer needed. It should be noted that since the future regional weather is directly related to the current regional weather, we design a heavier branch for the region and a lighter branch for the globe. In this way, DNN can focus on region-specific information, while the global weather state can also be introduced for guiding learning at a small computational cost.

Specifically, we deploy two independent input branches, i.e. heavy regional weather branch and light global weather branch. Each branch employs a patch embedding module[2], which consists of several convolution layers, to extract overlapping patches from regional weather $R_t \in \mathbb{R}^{H_r \times W_r \times D}$ and global weather $G_t \in \mathbb{R}^{H_g \times W_g \times D}$ respectively. These patches are then mapped into a C-dimensional embedding space. After passing through the patch embedding module, the global and regional weather data are represented by two features F_r^0 and $F_g^0 \in \mathbb{R}^{H \times W \times C}$. Additionally, we add position embedding $E_{pos} \in \mathbb{R}^{H \times W}$ to the regional weather feature, $F_r^0 = F_r^0 + E_{pos}$. Then, F_r^0 and F_g^0 are used for subsequent modeling.

3.2 Multi-Stage Regional-Global Feature Attention Mechanism

As discussed before, to achieve more reliable and accurate RWF, both regional weather and global weather state need to be considered simultaneously. Therefore, after obtaining the initial regional weather feature F_r^0 and global weather feature F_g^0, it is necessary to combine the two features for deep modeling. On the one hand, it requires making full use of the internal state of regional weather for learning in the heavy regional weather branch. On the other hand, global weather feature needs to be introduced at multiple stages of regional branch to fully integrate global weather information. To this end, as shown in Fig. 3 we propose a multi-stage regional-global feature attention mechanism.

Local Attention To effectively model regional weather, we focus on capturing localized effects within the region. This is crucial because regional weather

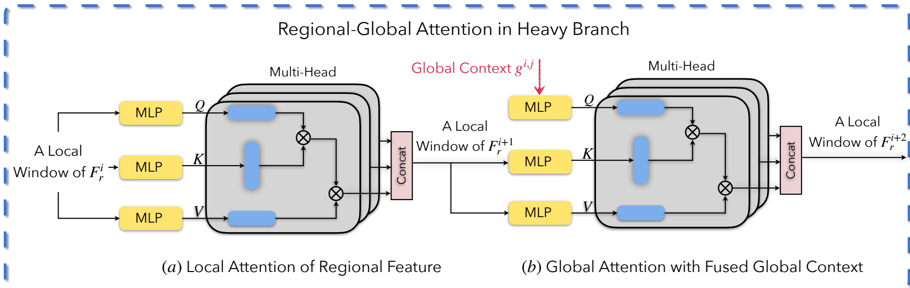

Fig. 3. Illustration of local attention and global attention in the proposed regional-global attention mechanism, where F_r^i is the regional weather feature.

is basically shaped by specific and fine-grained dynamics. To model localized patterns, we utilize local window attention [7], which first divides the regional weather feature F_r^0 into several non-overlapping windows and then applies attention within each window. By concentrating on individual windows, DNN learns specific local patterns in each window area more effectively. Furthermore, local window attention reduces the overall computational burden by limiting the scope of attention operations.

Specifically, we feed the regional weather feature F_r^0 into consecutive identical blocks of our model. As the input and output of each block share the same size, the calculation of the i-th block is as follows:

$$
\begin{aligned}
\hat{F}_r^i &= \text{W-MSA}\left(\text{LN}\left(F_r^{i-1}\right)\right) + F_r^{i-1}, \\
F_r^i &= \text{MLP}\left(\text{LN}\left(\hat{F}_r^i\right)\right) + \hat{F}_r^i
\end{aligned}
\tag{1}
$$

where F_r^{i-1} and F_r^i represent the input and output of i-th block respectively, W-MSA represents window multi-head self-attention [7], MLP stands for multi-layer perceptron, and LN is the layer normalization. It is notable that we only perform attention within each local window of the region up to now, and there is no interaction between different local windows.

Global Attention with Fused Global Context To realize global weather guidance for regional weather modeling, we aim to produce a global context feature, which can encompass global information and be incorporated into the regional weather branch. However, local attention based on local windows does not consider the global information of the region. To make full use of the two types of global information, we propose a light weight global context generator, which aims to merge the regional F_r^i and global weather feature F_g^j to generate a fused global context $g^{i,j} \in \mathbb{R}^{h_j \times w_j \times C}$.

The global context $g^{i,j}$ is subsequently integrated into the regional weather branch to guide the evolution of regional weather. This is achieved by customizing W-MSA into window multi-head global attention (W-MGA). Unlike W-MSA that uses the internal features of the local window as the query to calculate the attention weights, W-MGA uses the global context as the query to incorporate global information:

$$\hat{F}_r^i = \text{W-MGA}\left(g^{i,j}, \text{LN}\left(F_r^{i-1}\right)\right) + F_r^{i-1},$$
$$F_r^i = \text{MLP}\left(\text{LN}\left(\hat{F}_r^i\right)\right) + \hat{F}_r^i \tag{2}$$

After multi-stage feature extraction, we input the learned features to decoding head to yield. the forecast $\tilde{R}_{t+\Delta t} \in \mathbb{R}^{H_r \times W_r \times D}$.

3.3 Training Loss

To train our DNN model, we follow previous works [13, 9] and use the latitude-weighted mean squared error (MSE) as the loss function:

$$\mathcal{L} = \sum_{k=1}^{H_r}\sum_{i=1}^{W_r}\sum_{j=1}^{D} w(\phi_k) \cdot \left(R_{t+\Delta t}^{k,i,j} - \tilde{R}_{t+\Delta t}^{k,i,j}\right)^2, w(\phi_k) = \frac{\cos(\phi_k)}{\sum_{k=1}^{H_r}\cos(\phi_k)} \tag{3}$$

4 Experiment

4.1 Datasets

We use the data with 5.625° resolution from preprocessed ERA5 as global weather data. And follow the experiment of previous works [9], crop the NA region from the higher-resolution (i.e. 1.5°) global data for regional weather data.As to the variables in weather data, we focus on five key upper-air variables: geopotential (Z), specific humidity (Q), temperature (T), and the u and v components of wind speed (U and V), measured across 13 pressure levels. Additionally, we include four surface variables: 2m temperature (T2m), mean sea level pressure (MSLP), and the u and v components of 10m wind speed (U10, V10). We utilize 37 years of data (1979-2015) for training, one year of data (2016) for validation, and two years of data (2017-2018) for testing.

4.2 Evaluation Metrics and Baselines

To evaluate the performance, we adopt the commonly-used metrics, i.e. latitude-weighted RMSE and anomaly correlation coefficient (ACC):

$$\text{RMSE} = \sqrt{\sum_{k=1}^{H_r}\sum_{i=1}^{W_r}\sum_{j=1}^{D} w(\phi_k) \cdot \left(R_{t+\Delta t}^{k,i,j} - \tilde{R}_{t+\Delta t}^{k,i,j}\right)^2} \tag{4}$$

$$\text{ACC} = \frac{\sum_{k=1}^{H_r}\sum_{i=1}^{W_r}\sum_{j=1}^{D} w(\phi_k)(\tilde{R}_{t+\Delta t}^{k,i,j} - C^{k,i,j})(R_{t+\Delta t}^{k,i,j} - C^{k,i,j})}{\sqrt{\sum_{k,i,j} w(\phi_k)(\tilde{R}_{t+\Delta t}^{k,i,j} - C^{k,i,j})^2 \sum_{k,i,j} w(\phi_k)(R_{t+\Delta t}^{k,i,j} - C^{k,i,j})^2}} \tag{5}$$

where climatology C is the temporal mean of the ground truth data from 1979 to 2015. For RMSE, the lower the better. As for ACC, the higher the better.

Forecasts are evaluated at lead times ranging from 6 to 72 hours ($\Delta t = 6, 12, 18, 24, 36, 48, 72$), as many existing works focus on testing the performance of RWF methods for short-term forecasts up to 3 days [3].

We compare GCGRWF with the following methods: **ResNet** [4], **UNet** [14], **ClimaX** [9], **FourCastNet** [5]. All the above methods are trained from scratch on ERA5-NA dataset. We adopt the same training strategy for all methods.

4.3 Main Results

For simplicity, we follow previous works [1, 15, 5] to report results for 5 key weather variables: T850, Z500, T2m, U10, and MSLP. These variables are commonly used to assess the forecasting capability of RWF methods.

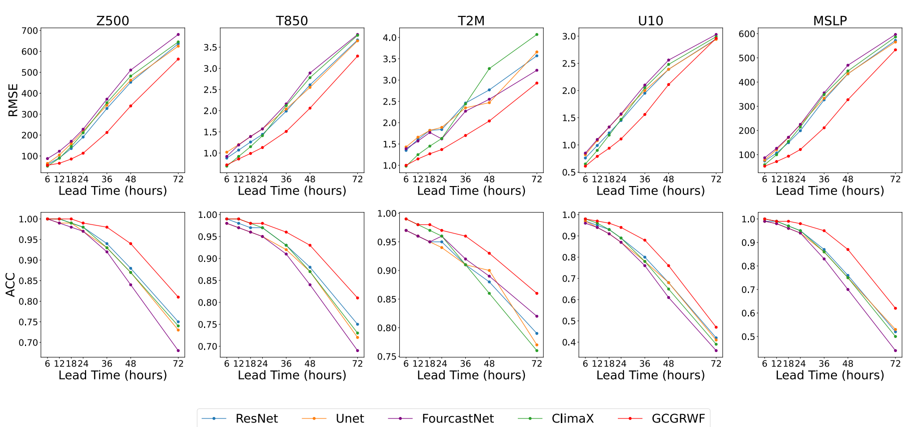

Fig. 4. RMSE (↓) and ACC (↑) comparison between our GCGRWF and other methods.

Experimental results are presented in Fig. 4, and we can draw the following conclusions: In terms of both RMSE and ACC, our GCGRWF is superior to other methods across most variables and lead times by a significant margin. As the lead time increases, performance improvement becomes more pronounced.

4.4 Ablation Study

To demonstrate the necessity and superiority of incorporating the global weather information for RWF, we compare the proposed GCGRWF with its basic variant GCGRWF-B, which does not use global weather data and is trained solely on ERA5-NA regional data. As shown in Fig. 5, GCGRWF consistently outperforms

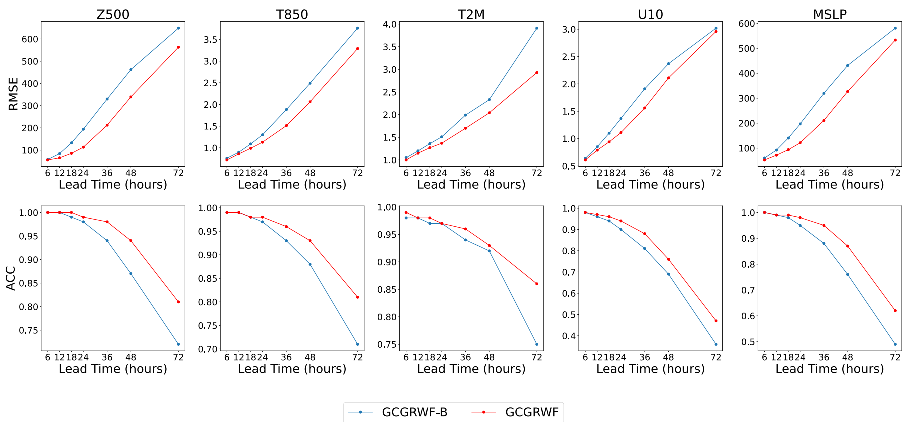

Fig. 5. Ablation study for the proposed GCGRWF.

GCGRWF-B across all lead times in terms of RMSE and ACC metrics, which verifies that considering global weather is beneficial for RWF. As the lead time increases, the performance gap between GCGRWF-B and GCGRWF gradually widens for most cases, which further indicates the importance of considering global weather information for RWF with larger lead time.

5 Conclusion

To fully consider external weather influence and avoid extra regional boundary processing for effective RWF, we propose a data-driven global context guided RWF method named GCGRWF. The key of our method is to directly utilize both regional weather data and global weather data to train a dual-branch DNN to predict future regional weather without any additional processing. Extensive experiments validate the effectiveness of our method and the significance of integrating global weather for RWF.

6 Acknowledgements

This research is partially supported by National Natural Science Foundation of China (Grant No. 62372460), National Natural Science Foundation of China (Grant No. 62402512), Hunan Provincial Natural Science Foundation of China (2024JJ4042), the science and technology innovation Program of Hunan Province (2024RC3134) and Youth Independent Innovation Science Fund of the National University of Defense Technology (ZK24-53).

References

1. Bi, K., Xie, L.e.: Accurate medium-range global weather forecasting with 3d neural networks. Nature **619**(7970), 533–538 (2023)
2. Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
3. Frnda, J., Durica, M., Rozhon, J., Vojtekova, M., Nedoma, J., Martinek, R.: Ecmwf short-term prediction accuracy improvement by deep learning. Scientific Reports **12**(1), 7898 (2022)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015), https://arxiv.org/abs/1512.03385
5. Kurth, T., et.al, S.: Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. In: Proceedings of the platform for advanced scientific computing conference. pp. 1–11 (2023)
6. Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al.: Learning skillful medium-range global weather forecasting. Science **382**(6677), 1416–1421 (2023)
7. Liu, Z., Lin, Y.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10012–10022 (2021)
8. Marbaix, P., Gallée, H., Brasseur, O., van Ypersele, J.P.: Lateral boundary conditions in regional climate models: a detailed study of the relaxation procedure. Monthly weather review **131**(3), 461–479 (2003)
9. Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J.K., Grover, A.: Climax: A foundation model for weather and climate. arXiv preprint arXiv:2301.10343 (2023)
10. Oskarsson, J., Landelius, T., Lindsten, F.: Graph-based neural weather prediction for limited area modeling. arXiv preprint arXiv:2309.17370 (2023)
11. Pu, Z., Kalnay, E.: Numerical weather prediction basics: Models, numerical methods, and data assimilation. Handbook of hydrometeorological ensemble forecasting pp. 67–97 (2019)
12. Qin, H., Chen, Y., Jiang, Q., Sun, P., Ye, X., Lin, C.: Metmamba: Regional weather forecasting with spatial-temporal mamba model. arXiv preprint arXiv:2408.06400 (2024)
13. Rasp, S., Dueben, P.D., Scher, S., Weyn, J.A., Mouatadid, S., Thuerey, N.: Weatherbench: a benchmark data set for data-driven weather forecasting. Journal of Advances in Modeling Earth Systems **12**(11), e2020MS002203 (2020)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation (2015), https://arxiv.org/abs/1505.04597
15. Xu, P., et.al, G.: Yinglong: Skillful high resolution regional short term forecasting with boundary smoothing. arXiv preprint arXiv:2401.16254 (2024)
16. Yáñez-Morroni, G., Gironás, J., Caneo, M., Delgado, R., Garreaud, R.: Using the weather research and forecasting (wrf) model for precipitation forecasting in an andean region with complex topography. Atmosphere **9**(8), 304 (2018)
17. Yuan, X., et.al, K.: The interconnected global climate system-a review of tropical–polar teleconnections. Journal of Climate **31**(15), 5765–5792 (2018)
18. Zhao, J., Zhan, R., Wang, Y., Xie, S.P., Zhang, L., Xu, M.: Lapsed el niño impact on atlantic and northwest pacific tropical cyclone activity in 2023. Nature Communications **15**(1), 6706 (2024)