Article:
Wang, Y., Zhao, B., Huang, P. et al. (2026) Improving medium-range temperature forecast over the Tibetan plateau through spatially adaptive fusion. Geophysical Research Letters, 53 (10). e2025GL121406. ISSN: 0094-8276

https://doi.org/10.1029/2025gl121406

Geophysical Research Letters

RESEARCH LETTER
10.1029/2025GL121406

Improving Medium-Range Temperature Forecast Over the Tibetan Plateau Through Spatially Adaptive Fusion

Yanfeng Wang[1,2], Bowen Zhao[3], Ping Huang[1,2], Lin Wang[1,2], Haosu Tang[4], Shuhao Ge[1,2,5], Mingwei Lu[1,2,5], Jingling Tang[1,2,5], and Yupeng Teng[6]

[1]State Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, [2]Center for Monsoon System Research, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, [3]Shanghai Typhoon Institute, China Meteorological Administration, Shanghai, China, [4]School of Geography and Planning, University of Sheffield, Sheffield, UK, [5]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China, [6]Meteorological Observation Centre, China Meteorological Administration, Beijing, China

Key Points:

- Swin Transformer fusion (STF) introduces learnable position encoding to improve 2-m temperature forecast over the Tibetan Plateau
- Compared with multi-model ensemble mean, STF reduces forecast biases by 29.14%–38.45% at 1–10-day lead
- STF achieves the greatest forecast improvements in high-bias regions through spatially adaptive fusion

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

P. Huang,
huangping@mail.iap.ac.cn

Citation:

Wang, Y., Zhao, B., Huang, P., Wang, L., Tang, H., Ge, S., et al. (2026). Improving medium-range temperature forecast over the Tibetan Plateau through spatially adaptive Fusion. *Geophysical Research Letters*, *53*, e2025GL121406. https://doi.org/10.1029/2025GL121406

Received 22 DEC 2025
Accepted 22 APR 2026

Abstract Tibetan Plateau exerts profound impacts on the global weather system, whereas its medium-range temperature forecast is challenging due to the complex topography. This study introduces Swin Transformer Fusion (STF), a spatially adaptive ensemble strategy that integrates forecasts from numerical weather prediction models (EC, GFS) and large meteorological models (Pangu, Fengwu). STF reduces 2-m temperature (T2m) forecast bias by 29.14%–38.45% across 1–10 days lead, outperforming conventional multi-model ensemble mean especially in high-bias regions. Attribution analysis reveals that Fengwu contributes most to STF's output despite Pangu's superior accuracy, challenging the assumption that higher-performing models should be weighed more heavily. These findings demonstrate that STF's spatially adaptive fusion enables region-specific integration of model strengths, offering an effective ensemble forecast strategy for topographically complex regions.

Plain Language Summary Predicting 2-m temperature (T2m) over the Tibetan Plateau remains challenging due to its complex topography. This study introduces a new deep learning–based method called Swin Transformer Fusion (STF), which improves medium-range T2m forecasts by combining predictions from multiple numerical weather prediction models (EC, GFS) and large meteorological models (Fengwu, Pangu). STF uses a position encoding scheme and spatially adaptive fusion to better capture spatial forecast characteristics. Compared with the multi-model ensemble mean, STF significantly reduces T2m forecast biases by 29.14%–38.45% at 1–10-day lead, especially in the original high-bias areas of Tibetan Plateau. Interestingly, although Pangu shows the best individual performance, Fengwu contributes most to STF's ensemble forecast. This challenges the common assumption that better-performing models should be weighted more in ensemble forecasting. By adaptively integrating model strengths across space, STF offers a new strategy for improving T2m forecasts in topographically complex regions like the Tibetan Plateau, with potential applications in other mountainous or data-sparse areas.

1. Introduction

The Tibetan Plateau, known as the Third Pole and Asia's Water Tower, exerts profound impacts on the global weather and climate system (He et al., 2025; T. Liu et al., 2023; G. Wu et al., 2023). Its high elevation induces unique dynamic and thermodynamic effects, which can influence East and South Asian monsoons and regulate the global energy and water cycle through water vapor transport (He et al., 2025; Q. Li et al., 2025; Sheng et al., 2022; S. Zhang et al., 2024). Central to these processes is the variability of surface air temperature (2-m temperature, T2m). This variability not only modulates the "elevated heat pump" effect but also acts as a precursor for extreme weather events through Rossby wave propagation (T. Liu et al., 2023; H. Tang et al., 2022; G. Wu et al., 2023; Yu et al., 2021; M. Zhao et al., 2025). Accurate medium-range T2m forecasts over the Tibetan Plateau are therefore critical for mitigating both local hazards (e.g., snowmelt-induced hydrological extremes) and remote climate impacts, including East Asian heatwaves and South Asian floods (He et al., 2025; T. Liu et al., 2023; Sheng et al., 2022; Sun et al., 2023; Yang et al., 2025).

Over recent decades, numerical weather prediction (NWP) models have achieved notable improvements in T2m forecasting over the Tibetan Plateau (Bougeault et al., 2010; He et al., 2025; G. Wu et al., 2023; Yang et al., 2025;

H. Zhang et al., 2018). The improvement of forecast skill is mainly driven by technological progress, such as advancements in multi-source data assimilation (satellite and ground station integration) and refined parameterization of cryosphere-atmosphere interactions over the Tibetan Plateau (He et al., 2025; G. Wu et al., 2023; Yang et al., 2025). The emergence of large meteorological models, enabled by supercomputing and the proliferation of Earth system data, has further enhanced medium-range forecast skill (Bi et al., 2023; Chen et al., 2023; Hersbach et al., 2020; P. Lin et al., 2022; C.-C. Liu et al., 2024). Despite these advances, both NWP and large meteorological models struggle with forecast challenges from complex topography, resulting in substantially higher forecast biases over the Tibetan Plateau compared to East Asian plains (Figure S1 in Supporting Information S1) (Y. Tang et al., 2022; G. Wu et al., 2023; Yang et al., 2025; S. Zhang et al., 2024).

Multimodel ensemble mean (MME) is a widely used strategy to improve forecast skill by averaging across models with opposing biases (Barnston et al., 2019; Bi et al., 2023; Bougeault et al., 2010; Chen et al., 2023; C.-C. Liu et al., 2024; Y. F. Wang et al., 2020). However, MME performance faces two key limitations: (a) common systematic biases across models may be amplified rather than offset and (b) inclusion of low-skill members can degrade ensemble performance relative to the best single model (Barnston et al., 2019; Ding et al., 2020; Gao et al., 2021). While weighted ensemble methods partially address these limitations by assigning higher weight to skillful models, their reliance on localized linear assumptions fails to capture nonlinear relationships and remote signals (Gao et al., 2021; Sloughter et al., 2013; H. Wang et al., 2024; T. Wang & Huang, 2024). In light of these limitations, there is a growing demand for more adaptive weighting strategies that can capture multi-scale spatial dependencies and nonlinear relationships, particularly in high-bias regions such as the Tibetan Plateau.

Deep learning has emerged as a promising tool for post-processing multi-model forecasts (Cui et al., 2025; Tian et al., 2024; Tong & Zhou, 2024; Xia & Wang, 2025; Y. Zhang et al., 2024), offering powerful capabilities for capturing complex spatial patterns and nonlinear relationships (Huang et al., 2025; Z. Liu et al., 2023; Qin et al., 2024; B. Wang et al., 2024; B. Zhao et al., 2025; L. Zhou & Zhang, 2022, 2023). Previous studies have demonstrated that combining the strengths of large meteorological models or using deep learning to correct NWP outputs, can improve forecast performance of NWP. This study proposes a novel application of Swin Transformer to perform fusion-based correction of multi-model forecasts (Z. Liu et al., 2021), significantly improving the medium-range T2m forecast skill over the Tibetan Plateau. Swin Transformer fusion (STF) can make use of the unique position encoding scheme and shifted window architecture to extract multi-scale spatial features (Z. Liu et al., 2021). It can capture localized weather signals through non-overlapping window operations while preserving remote weather signals via shifted window connections. We also compare the performance of STF with MME and U-Net (H. Lin et al., 2023) in this paper to demonstrate the superiority of STF.

2. Data and Methods

2.1. Reanalysis and Model Data

In this study, we use the 10-day forecast of T2m from two widely used NWP models, the European Centre for Medium-Range Weather Forecasts (hereafter denoted as EC) (Bougeault et al., 2010) and Global Ensemble Forecast System (GFS) from National Centers for Environmental Prediction (X. Zhou et al., 2022) and two large meteorological models, Pangu-Weather (Pangu) (Bi et al., 2023) and Fengwu (Chen et al., 2023). Analysis focuses on daily T2m predictions from 2021 to 2023 at 0.25° × 0.25° resolution over the Tibetan Plateau (26.25°N–42°N, 73°E–104.75°E). In this study, all models are initialized daily at 00:00 UTC and produce forecasts for lead time ranging from 1 to 10 days. These multi-model daily predictions serve as predictors to the STF (Z. Liu et al., 2021), with ERA5 reanalysis as the target (Hersbach et al., 2020). First, EC and GFS are chosen because they are two of the most widely used NWP models, and their data is publicly available. Second, Pangu and Fengwu are selected as they are representative large meteorological models that are computationally feasible within our existing environment. The inference for these large meteorological models is performed on Deep Computing Units (DCUs, 16–32 GB of video memory) within the EarthLab supercomputing platform. Given this hardware limitation, we are currently unable to run inference on larger meteorological models. The model training period spans 20210101–20230930, with independent validation for 20231001–20231130 and test for 20231201–20231231. In 20210101–20231231, spatial distribution of T2m forecast bias vary significantly among these four models. While the GFS model's biases are prominent in both the northwest and southeast of the Tibetan Plateau, the other three models have their largest errors confined to the southeast (Figure S1 in Supporting Information S1).

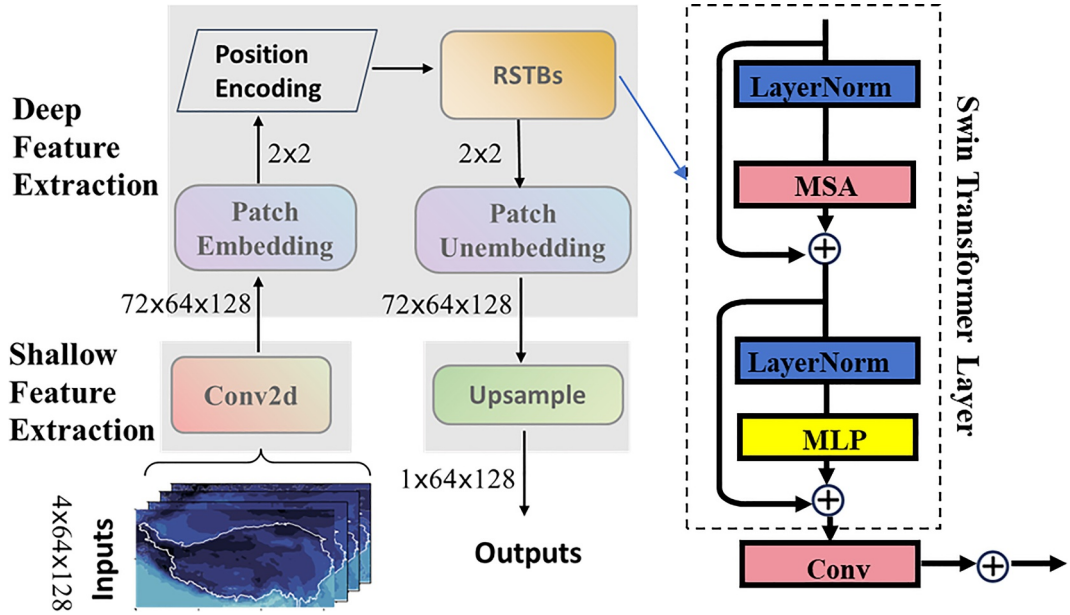

Figure 1. Structure of the Swin Transformer fusion model for ensemble forecast of T2m.

2.2. Methods

STF makes use of the hierarchical design and shifted window architecture to integrate predictions from EC, GFS, Pangu, and Fengwu. We design a specific channel-wise concatenation scheme to integrate the outputs from NWP and large meteorological models. The novelty of our design lies not in the operation itself, but in the specific formulation of the input tensor to fuse heterogeneous forecasting from both NWP and large meteorological models. By treating these distinct sources as separate channels, we allow the network to explicitly learn the complementary correlations between NWP and large meteorological models, which is not achievable through simple averaging. The batch size is 16 and the epochs are 100. We use the Adam optimizer with the learning rate at 0.0001.

STF contains a three-stage hierarchical architecture (Figure 1):

1. Shallow Feature Extraction Module: The initial processing stage uses a single 3×3 convolutional layer to map input data into an embedded feature space, which expand the channels from 4 to 72;
2. Deep Feature Extraction Module: This module contains several key components. First, Patch Embedding divides the input data into non-overlapping 2×2 pixel patches, which are then flattened into feature vectors. These vectors are then fed into the core of the module: the Residual Swin Transformer Blocks (RSTBs). This component has four stages, with each stage being an RSTB containing 6 Swin Transformer Layers and one convolution layer. The structure of the Swin Transformer Layer is shown in the right part of Figure 1, with a normalization layer (LayerNorm), a multi-head self-attention layer (MSA), a fully connected layer (MLP), and residual connections represented by plus signs. As the fundamental component for feature extraction, these RSTBs are responsible for progressively learning rich and hierarchical feature representations from the input data; then Patch Unembedding will reconstruct the patches back into a continuous spatial grid
3. Upsample Module: This module employs adaptive pixel-shuffle operations with parameterized convolution kernels to achieve resolution matching with the original input data.

Here, we specially adopt a learnable absolute position encoding scheme. For the input feature map X, we initialize a trainable position encoding matrix P, which is added to input features and automatically learned through backpropagation:

$$X' = X + P \tag{1}$$

For the evaluation of STF, we use the root mean square error (RMSE) metric, which is weighted by latitude to account for the varying spatial scales at different latitudes (Bi et al., 2023; Chen et al., 2023). The weighted RMSE is expressed as follows:

$$\text{weighed RMSE(lead)} = \sqrt{\frac{\sum_{i=1}^{n\text{lat}} \sum_{j=1}^{n\text{lon}} L(i)\left(\text{prediction}_{i,j,\text{lead}} - \text{Obs}_{i,j,\text{lead}}\right)^2}{n\text{lat} \times n\text{lon}}} \tag{2}$$

$$L(i) = \frac{\cos\theta_i}{\sum_{i=1}^{n\text{lat}} cos\theta_i} \tag{3}$$

where i and j are the latitude and longitude indices, respectively. $\text{Obs}_{i,j,\text{lead}}$ represents the observed value, $\text{prediction}_{i,j,\text{lead}}$ is the corresponding predicted value. $L(i)$ is the weight at latitude θ_i. The $n$lat and $n$lon are the total number of data points across latitudes and longitudes. The latitudinal weights are calculated to give more weight to the equatorial regions, where the physical distance between grid points is greater than that in the polar regions.

We use an explainable artificial intelligence (AI) method, integrated gradients, to evaluate the contribution of each predictor to the final fusion prediction results (Kapishnikov et al., 2021; Sundararajan et al., 2017; H. Wang et al., 2024), calculated as:

$$\text{Integrated gradients}(i,j) = (x_{i,j} - x_{\text{ref}}) \int \frac{\partial \text{STF}(x_{\text{ref}} + \alpha(x_{i,j} - x_{\text{ref}}))}{\partial x_{i,j}} \text{d}\alpha \tag{4}$$

where x_{ref} is the baseline, usually set to zero; $x_{i,j}$ is the predictor value at grid (i, j); STF is the Swin Transformer fusion model.

3. Results

3.1. Evaluation of T2m Forecast Skill Over the Tibetan Plateau

Figure 2 presents an evaluation of T2m forecast skill over the Tibetan Plateau from the individual models EC, GFS, Fengwu, Pangu, and two ensemble strategies, MME and STF across 1–10-day lead during the test period (20231201–20231231). Among the single models, Pangu consistently delivers the highest forecast accuracy, followed by Fengwu, EC, and GFS (Figure 2a). Notably, large meteorological models (Fengwu, Pangu) outperform the NWP models (EC, GFS) across all lead time, suggesting that data-driven architectures may better capture nonlinear processes over complex terrain (Figure 2a and Figure S2 in Supporting Information S1). However, MME, which linearly averages the four single-model forecasts, fails to surpass either Fengwu or Pangu, underscoring the limitations of static weighting in regions with persistent forecast biases. In contrast, STF demonstrates a substantial advancement in ensemble forecasting, outperforming all single models and the MME across the entire forecast horizon (Figure 2a, black line). STF reduces the RMSE between forecast and observation by 29.14% (9-day lead) to 38.45% (1-day lead) relative to MME, and by 9.56% (1-day lead) to 26.41% (10-day lead) compared to the best-performing Pangu. Importantly, the magnitude of RMSE reduction increases with lead time.

To validate STF's superior ability to capture multi-scale spatiotemporal dependencies, we also conduct a sensitive experiment comparing STF with and without the learnable position encoding scheme in Figure S2 in Supporting Information S1. After removing the learnable position encoding scheme of STF, the T2m forecasting bias across 1–10-day lead significantly increases by 3.91%–12.32%. We have also conducted an ablation study on the RSTBs (Figure S2 in Supporting Information S1). By removing RSTBs which are responsible for processing features at different scales, we observed a significant drop in forecasting skill. The forecasting bias will increase by 3.48%–26.70% across 1–10-day lead. We also add the ensemble forecasting results based on U-Net (Figure S2 in Supporting Information S1) to further demonstrate STF's advantages in ensemble forecasting. Compared with STF, U-Net based ensemble forecasting will increase the forecasting bias by 3.00%–17.26% at 1–10-day lead. Furthermore, spatial distribution of STF with and without positional encoding in Figure S3 in Supporting Information S1 also demonstrates that the inclusion of positional encoding leads to better performance and suggests a promising ensemble forecasting in high-bias, topographically complex regions.

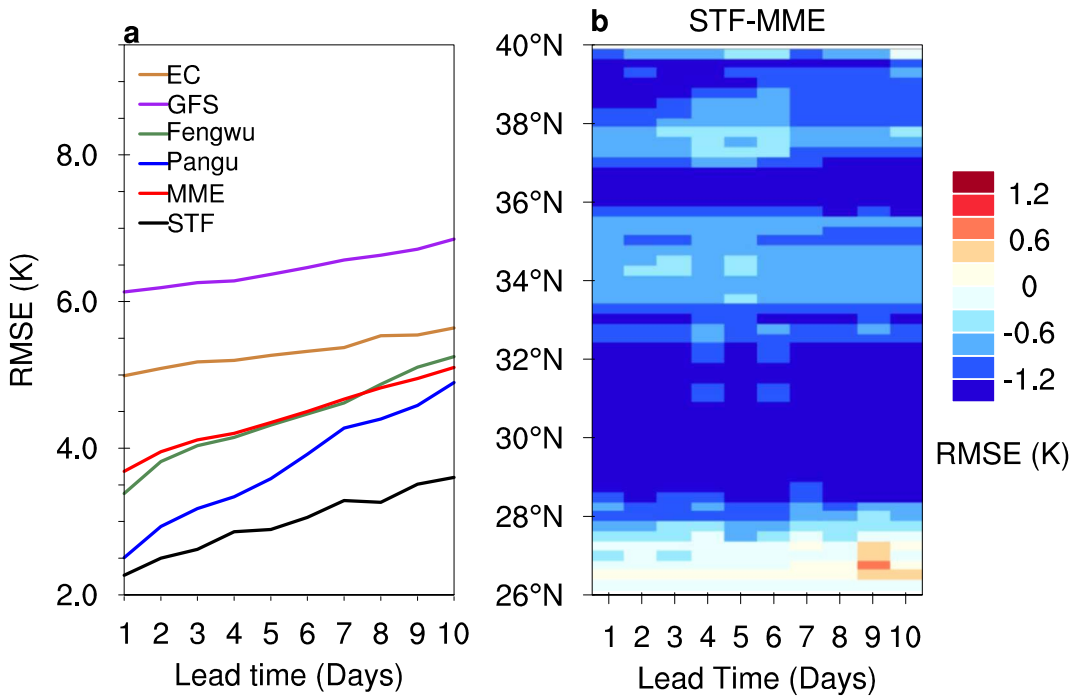

Figure 2. (a) RMSE (unit: K) between ERA5 and T2m forecast at 1–10-day lead from EC, Global Ensemble Forecast System, Pangu, Fengwu, MME, and STF in 20231201–20231231; (b) relative reduction of T2m forecast in RMSE for STF compared to the MME across various latitudes from 20231201 to 20231231.

To further assess spatial forecast performance, we analyzed the differences in zonal-mean RMSE between STF and MME, each relative to observations across the Tibetan Plateau (Figure 2b). STF consistently outperforms MME across nearly all latitudes and lead time, demonstrating robust spatial generalization. Notably, the improvement is most pronounced in the high-altitude belt (28°N–36°N, $P < 0.1$), a region historically associated with elevated forecast uncertainty due to complex topography and sparse observations. These results underscore STF's potential not only for regional-scale forecast improvement but also for grid-point correction, offering a scalable solution for ensemble post-processing in topographically complex regions.

3.2. Spatial Feature of the Improved T2m Forecast Skill From STF

To evaluate the spatial distribution of improved forecast skill achieved by STF, we computed the spatial RMSE for four single models (EC, GFS, Fengwu, and Pangu) and MME, along with the RMSE reduction achieved by STF at 10-day lead (Figure 3). All single models and MME exhibit pronounced forecast biases in the high-altitude Tibetan Plateau (28°N–36°N, 80°E–100°E), with relatively smaller biases in 36°N–42°N. Specifically, GFS shows an RMSE > 5 K across most Tibetan Plateau regions (Figure 3c), while EC and Fengwu display peak biases (RMSE > 5 K) in the southeastern Tibetan Plateau near the Hengduan Mountains (Figures 3a and 3e). Pangu and MME exhibit high biases (RMSE > 5 K) in both southeastern and southwestern Tibetan Plateau, although their biases in the Hengduan Mountains are significantly smaller than those of EC, GFS, and Fengwu (Figures 3h and 3j).

STF demonstrates a targeted and systematic reduction of forecast bias, leveraging model diversity to improve forecast skill across nearly all grid points. Compared to GFS, STF reduces RMSE by over 3K across the entire Tibetan Plateau (Figure 3d), and delivers the most substantial improvements for EC and Fengwu in the Hengduan Mountains (Figures 3b and 3f). Compared with Pangu, STF also effectively reduces T2m forecast biases, particularly in the southwestern plateau. Importantly, the spatial distribution of STF's improvements (Figure 3, right column) closely mirrors the original RMSE patterns of the single models and MME (Figure 3, left column), suggesting that STF's training process is explicitly sensitive to high-bias regions.

10.1029/2025GL121406

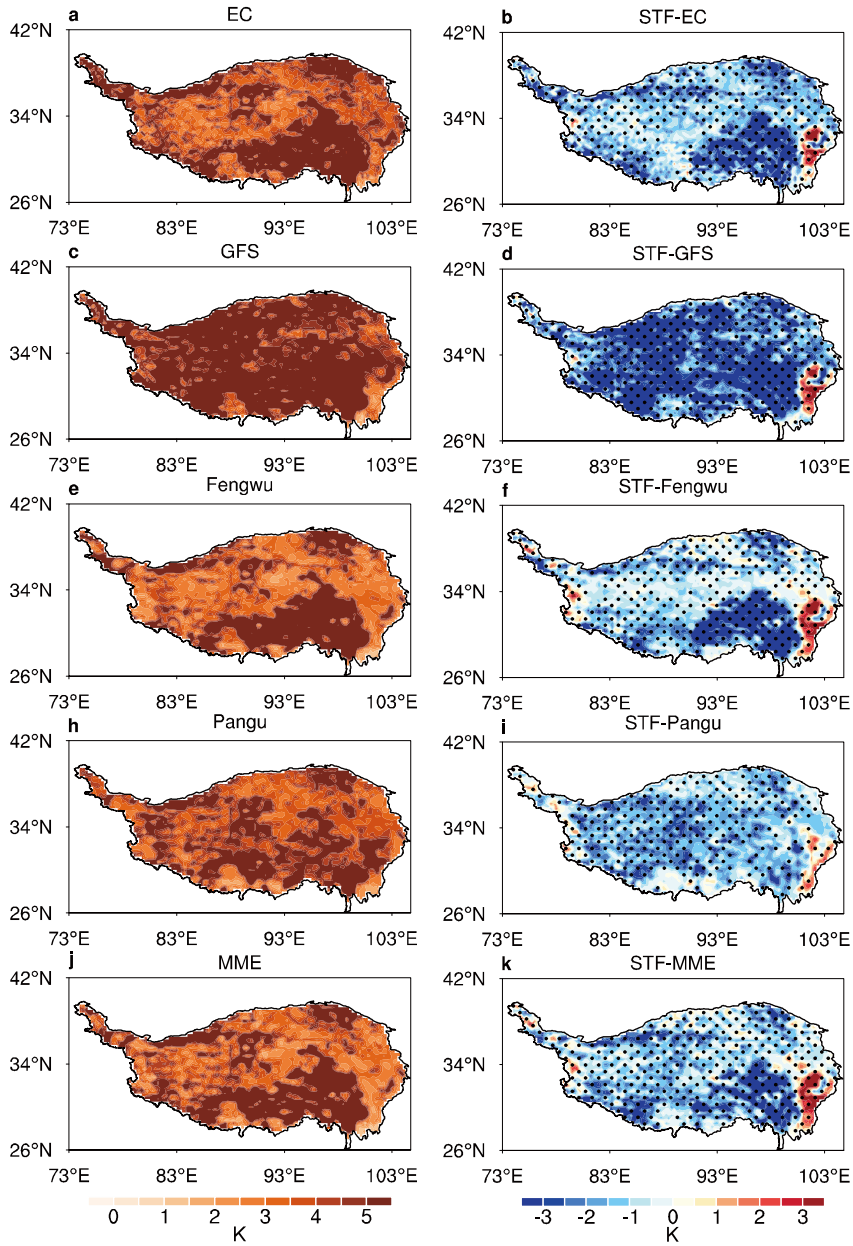

Figure 3. (a, c, e, h, and j) Spatial RMSE of T2m forecast (unit: K) at 10-day lead time, showcasing the performance of EC, GFS, Fengwu, Pangu, and MME in 20231201–20231231; (b, d, f, i, and k) The spatial distribution of STF's reduced forecast RMSE compared with EC, GFS, Fengwu, Pangu, and MME at 10-day lead time. The black dotted regions indicate areas where STF's performance improvement is statistically significant at the 0.1 level of the Student's *t* test.

This alignment is likely caused by STF's loss function design, which emphasizes bias reduction in high-RMSE grids during parameter optimization (Bi et al., 2023; Chen et al., 2023). Such behavior reflects a bias-aware learning mechanism, enabling STF to prioritize correction in high-bias region. STF's learnable position encoding enables the model to internalize spatial dependencies and implicitly capture topographic features, elevation gradients, and terrain-induced biases, without explicit terrain inputs. Rather than uniformly averaging model outputs, STF performs spatially adaptive fusion, integrating model-specific strengths in a topographically informed and dynamically responsive manner.

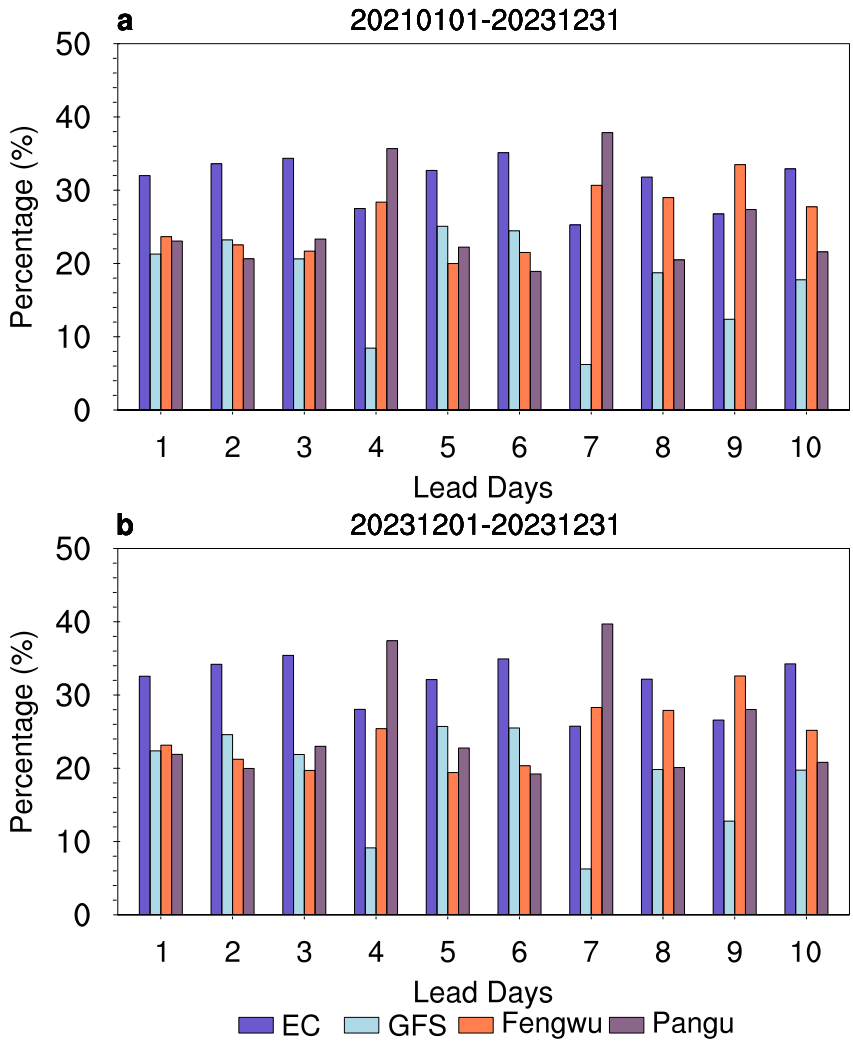

Figure 4. The contribution of EC, Global Ensemble Forecast System, Fengwu, and Pangu to the final Swin Transformer fusion forecast results at 1–10-day lead time in (a) 20210101–20231231 and (b) 20231201–20231231. Contribution is quantified by integrated gradients.

3.3. Attribution Analysis of STF's Forecast Composition

We apply integrated gradients across 1–10-day lead to quantify the contribution of each model to STF's ensemble forecast (Figure 4). Unlike MME's static averaging, STF exhibits dynamic and region-specific contribution patterns. STF exhibits distinct and dynamical contribution patterns. Across 1–10-day lead, large meteorological models (Fengwu and Pangu) contribute more to STF's forecasts than NWP models (EC and GFS) for both the entire period (20210101–20231231, Figure 4a) and the independent test period (20231201–20231231, Figure 4b). Meanwhile, GFS is the model with the lowest T2m forecast skill, but it is not the smallest contributor at 5-day lead.

These results highlight a key innovation of STF. Model contributions are not statically assigned based on performance metrics (Sloughter et al., 2013; H. Wang et al., 2024; T. Wang & Huang, 2024), but are instead dynamically inferred through spatially adaptive fusion (Z. Liu et al., 2021). STF's learnable position encoding collectively enable the model to identify and leverage region-specific advantages from each input source. For instance, Fengwu may offer superior spatial features in specific regions (notably in the northwest of the Tibetan Plateau), making it more influential in the final ensemble despite higher overall RMSE than Pangu.

Analysis of relative impact of local versus remote signals on the key Tibetan Plateau region (28°N–36°N, 80°E–100°E) shows that local weather signals dominate STF's prediction, accounting for over 90% of the ensemble contribution and peaking at 97.53% on the 7-day lead (Table S1 in Supporting Information S1). Remote signals captured via the shifted window architecture contribute modestly, reaching 6.60% at 3-day lead.

4. Conclusion and Discussion

This study demonstrates STF as a highly effective ensemble strategy for multi-model T2m forecasts over the Tibetan Plateau. By combining forecasts from NWP models (EC, GFS) and large meteorological models (Pangu, Fengwu), STF reduces forecast bias by 29.14%–38.45% across 1–10-day lead, significantly outperforming the conventional MME approach. Through spatial RMSE analysis, STF's learnable position encoding scheme and strategic error-targeting mechanism are shown to preferentially correct high-bias regions (especially in 93°-E–100°E, Figure 3) while maintaining superior overall skill compared to the conventional MME. The RMSE reduction patterns align closely with the original bias distributions of the single models, indicating effective targeting of bias sensitive zones.

Moreover, attribution analysis using integrated gradients reveals STF's departure from traditional ensemble methods that rely on static performance-based weighting. Instead, STF dynamically infers model contributions through spatially adaptive fusion, guided by its RSTBs and embedded positional encoding. Based on the ablation experiments, the forecasting bias will increase by 3.48%–26.70% at 1–10-day lead without RSTBs. And without the learnable position encoding scheme, the T2m forecasting bias will increase by 3.91%–12.32% at 1–10-day lead (20231201–20231231). Notably, although Pangu exhibits the highest overall forecast skill, Fengwu contributes most to STF's final output, underscoring STF's capacity to extract region-specific strengths rather than relying on overall rankings.

By leveraging model diversity and selectively integrating both localized and remote weather signals, STF offers a position-sensitive solution for ensemble post-processing in topographically complex regions. This study introduces a new ensemble strategy design principles, demonstrating that spatially adaptive fusion significantly outperforms conventional MME in bias mitigation and forecast accuracy. Besides, incorporating a broader range of models, such as CMA-MESO (H. Li et al., 2024; Z. Wu et al., 2025), is a probably valuable direction for improving the robustness of ensemble forecasting. We did consider adding other models like CMA-MESO (H. Li et al., 2024; Z. Wu et al., 2025). However, the CMA-MESO data is only publicly available for the most recent 7-day period, which is insufficient for the comprehensive analysis required in our study. In the future, if data availability and hardware permit, we will incorporate more models for ensemble forecasting.

Conflict of Interest

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

Availability Statement

All the data and models used in this paper are publicly available.

ERA5 (Hersbach et al., 2020): https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels.

EC (Bougeault et al., 2010): https://www.ecmwf.int/en/forecasts/datasets/set-iii.

GFS (X. Zhou et al., 2022): https://www.nco.ncep.noaa.gov/pmb/products/gfs/.

Code of Pangu (Bi et al., 2023): https://github.com/198808xc/Pangu-Weather.

Code of Fengwu (Chen et al., 2023): https://github.com/OpenEarthLab/FengWu.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (42425504, 42205035, and 42205145), the China Postdoctoral Science Foundation (2023T160632 and 2022M723096), and the CAS Project for Young Scientists in Basic Research (YSBR-137). We also thank the technical support of the National Large Scientific and Technological Infrastructure "Earth System Numerical Simulation Facility" (https://cstr.cn/31134.02.EL). We thank the two reviewers for their suggestions and comments, which have greatly improved the quality of the manuscript. We would also like to express our gratitude to Dr. Chen Sheng from the Institute of Atmospheric Physics for suggestions on Tibetan Plateau research and Dr. Lu Zhou from Nanjing University of Information Science and Technology for advice on integrated gradient calculations.

References

Barnston, A. G., Tippett, M. K., Ranganathan, M., & L'Heureux, M. L. (2019). Deterministic skill of ENSO predictions from the north American multimodel ensemble. *Climate Dynamics*, *53*(12), 7215–7234. https://doi.org/10.1007/s00382-017-3603-3

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, *619*(7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D. H., et al. (2010). The THORPEX interactive grand global ensemble. *Bulletin of the American Meteorological Society*, *91*(8), 1059–1072. https://doi.org/10.1175/2010BAMS2853.1

Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.-J., et al. (2023). FengWu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. https://doi.org/10.48550/arXiv.2304.02948

Cui, W., Si, J., Zhang, L., Han, L., & Chen, Y. (2025). Enhanced multimodal-fusion network for radar quantitative precipitation estimation incorporating relative humidity data. *IEEE Transactions on Geoscience and Remote Sensing*, *63*, 1–13. https://doi.org/10.1109/TGRS.2025.3585737

Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2020). Relating CMIP5 model biases to seasonal forecast skill in the tropical Pacific. *Geophysical Research Letters*, *47*(5), e2019GL086765. https://doi.org/10.1029/2019gl086765

Gao, Y., Wallach, D., Hasegawa, T., Tang, L., Zhang, R., Asseng, S., et al. (2021). Evaluation of crop model prediction and uncertainty using Bayesian parameter estimation and Bayesian model averaging. *Agricultural and Forest Meteorology*, *311*, 108686. https://doi.org/10.1016/j.agrformet.2021.108686

He, B., He, X., Liu, Y., Wu, G., Bao, Q., Hu, W., et al. (2025). Role of thermal and dynamical subdaily perturbations over the Tibetan Plateau in 30-day extended-range forecast of East Asian precipitation in early summer. *npj Climate and Atmospheric Science*, *8*(1), 40. https://doi.org/10.1038/s41612-025-00931-2

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Huang, Y., Zhou, Z.-Q., Zhang, R., Zhong, X., & Li, H. (2025). The impact of tropical convective heating on the Amundsen Sea low in a deep-learning weather forecasting model. *Advances in Atmospheric Sciences*, *42*(12), 2411–2421. https://doi.org/10.1007/s00376-025-4360-6

Kapishnikov, A., Venugopalan, S., Avci, B., Wedin, B., Terry, M., & Bolukbasi, T. (2021). Guided integrated gradients: An adaptive path method for removing noise. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (pp. 5048–5056).

Li, H., Yang, Y., Sun, J., Jiang, Y., Gan, R., & Xie, Q. (2024). A 3D-Var assimilation scheme for vertical velocity with CMA-MESO v5.0. *Geoscience Model Development*, *17*(15), 5883–5896. https://doi.org/10.5194/gmd-17-5883-2024

Li, Q., Xue, Y., Kong, X., Lau, W. K. M., Wang, A., Li, Q., et al. (2025). Excessive Tibetan Plateau spring warming found to cause catastrophic June 2024 heavy rainfall in China. *Scientific Bulletin*, *70*(10), 1596–1600. https://doi.org/10.1016/j.scib.2025.01.011

Lin, H., Tang, J., Wang, S., Wang, S., & Dong, G. (2023). Deep learning downscaled high-resolution daily near surface meteorological datasets over east Asia. *Scientific Data*, *10*(1), 890. https://doi.org/10.1038/s41597-023-02805-9

Lin, P., Zhao, B., Wei, J., Liu, H., Zhang, W., Chen, X., et al. (2022). The super-large ensemble experiments of CAS FGOALS-g3. *Advances in Atmospheric Sciences*, *39*(10), 1746–1765. https://doi.org/10.1007/s00376-022-1439-1

Liu, C.-C., Hsu, K., Peng, M. S., Chen, D.-S., Chang, P.-L., Hsiao, L.-F., et al. (2024). Evaluation of five global AI models for predicting weather in eastern Asia and western Pacific. *npj Climate and Atmospheric Science*, *7*(1), 221. https://doi.org/10.1038/s41612-024-00769-0

Liu, T., Chen, D., Yang, L., Meng, J., Wang, Z., Ludescher, J., et al. (2023). Teleconnections among tipping elements in the Earth system. *Nature Climate Change*, *13*(1), 67–74. https://doi.org/10.1038/s41558-022-01558-4

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *ICCV 2021*, 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986

Liu, Z., Zhou, W., & Yuan, Y. (2023). 3D DBSCAN detection and parameter sensitivity of the 2022 Yangtze river summertime heatwave and drought. *Atmospheric and Oceanic Science Letters*, *16*(4), 100324. https://doi.org/10.1016/j.aosl.2022.100324

Qin, B., Yang, Z. Y., Mu, M., Wei, Y. T., Cui, Y. H., Fang, X. H., et al. (2024). The first kind of predictability problem of El Niño predictions in a multivariate coupled data-driven model. *Quarterly Journal of the Royal Meteorological Society*, *150*(765), 5452–5471. https://doi.org/10.1002/qj.4882

Sheng, C., He, B., Wu, G., Liu, Y., & Zhang, S. (2022). Interannual influences of the surface potential vorticity forcing over the Tibetan Plateau on East Asian summer rainfall. *Advances in Atmospheric Sciences*, *39*(7), 1050–1061. https://doi.org/10.1007/s00376-021-1218-4

Sloughter, M., Gneiting, T., & Raftery, A. E. (2013). Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Monthly Weather Review*, *141*(6), 2107–2119. https://doi.org/10.1175/MWR-D-12-00002.1

Sun, W., Wu, G., Liu, Y., Mao, J., Zhuang, M., & Liu, X. (2023). Delayed response of the onset of the summer monsoon over the Bay of Bengal to land–sea thermal contrast. *Journal of Climate*, *36*(12), 4051–4070. https://doi.org/10.1175/JCLI-D-22-0478.1

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *Proceedings of the 34th international conference on machine learning* (pp. 3319–3328). JMLR.org.

Tang, H., Hu, K., Huang, G., Wang, Y., & Tao, W. (2022). Intensification and northward extension of northwest Pacific anomalous anticyclone in El Niño decaying mid-summer: An energetic perspective. *Climate Dynamics*, *58*(1), 591–606. https://doi.org/10.1007/s00382-021-05923-5

Tang, Y., Duan, A., Xiao, C., & Xin, Y. (2022). The prediction of the Tibetan Plateau thermal condition with machine learning and Shapley additive explanation. *Remote Sensing*, *14*(17), 4169. https://doi.org/10.3390/rs14174169

Tian, W., Yi, L., Niu, X., Fang, R., Zhang, L., Liu, H., et al. (2024). RadarNet: A parallel spatiotemporal encoder network for radar extrapolation. *Neurocomputing*, *591*, 127665. https://doi.org/10.1016/j.neucom.2024.127665

Tong, X., & Zhou, W. (2024). Assessing predictive attribution in NMME forecasts of summer precipitation in eastern China using deep learning. *npj Climate and Atmospheric Science*, *7*(1), 304. https://doi.org/10.1038/s41612-024-00835-7

Wang, B., Wei, W., Yin, Z., & Xu, L. (2024). Using machine learning to analyze the changes in extreme precipitation in southern China. *Atmospheric Research*, *302*, 107307. https://doi.org/10.1016/j.atmosres.2024.107307

Wang, H., Zheng, X.-T., Cai, W., Han, Z.-W., Xie, S.-P., Kang, S. M., et al. (2024). Atmosphere teleconnections from abatement of China aerosol emissions exacerbate northeast Pacific warm blob events. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(21), e2313797121. https://doi.org/10.1073/pnas.2313797121

Wang, T., & Huang, P. (2024). Superiority of a convolutional neural network model over dynamical models in predicting central Pacific ENSO. *Advances in Atmospheric Sciences*, *41*(1), 141–154. https://doi.org/10.1007/s00376-023-3001-1

Wang, Y. F., Zhang, Z. H., & Huang, P. (2020). An improved model-based analogue forecasting for the prediction of the tropical Indo-Pacific sea surface temperature in a coupled climate model. *International Journal of Climatology*, *40*(15), 6346–6360. https://doi.org/10.1002/joc.6584

Wu, G., Zhou, X., Xu, X., Huang, J., Duan, A., Yang, S., et al. (2023). An integrated research plan for the Tibetan Plateau land–air coupled system and its impacts on the global climate. *Bulletin of the American Meteorological Society*, *104*(1), E158–E177. https://doi.org/10.1175/BAMS-D-21-0293.1

Wu, Z., Han, W., Xie, H., Ye, M., & Gu, J. (2025). Assimilation of FY-3G Ku-band radar observations with 1D Bayesian retrieval and 3DVAR in CMA-MESO. *Quarterly Journal of the Royal Meteorological Society*, *151*(770), e4964. https://doi.org/10.1002/qj.4964

Xia, H., & Wang, K. (2025). Hourly, kilometer-scale precipitation merged from rain gauge, ground-based radar and satellites over east Asia: Methods, evaluation and applications. *Journal of Hydrology*, *662*, 134148. https://doi.org/10.1016/j.jhydrol.2025.134148

Yang, K., Chen, J., Zhao, T., Lu, C., Xu, X., Luo, Y., et al. (2025). Effects of fine terrain complexity on cloud and precipitation changes over the Tibetan Plateau: A modeling study. *npj Climate and Atmospheric Science*, *8*(1), 22. https://doi.org/10.1038/s41612-025-00907-2

Yu, W., Liu, Y., Yang, X.-Q., Wu, G., He, B., Li, J., & Bao, Q. (2021). Impact of north Atlantic SST and Tibetan Plateau forcing on seasonal transition of springtime South Asian monsoon circulation. *Climate Dynamics*, *56*(1), 559–579. https://doi.org/10.1007/s00382-020-05491-0

Zhang, H., Zhang, F., Zhang, G., Che, T., & Yan, W. (2018). How accurately can the air temperature lapse rate over the Tibetan Plateau be estimated from MODIS LSTs? *Journal of Geophysical Research: Atmospheres*, *123*(8), 3943–3960. https://doi.org/10.1002/2017JD028243

Zhang, S., Hu, Q., Meng, X., Lü, Y., & Yang, X. (2024). Spatiotemporal evaluation and future projection of diurnal temperature range over the Tibetan Plateau in CMIP6 models. *Advances in Atmospheric Sciences*, *41*(11), 2245–2258. https://doi.org/10.1007/s00376-024-3346-0

Zhang, Y., Geng, S., Ma, G., Zhu, L., & Liu, Q. (2024). An improvement multitask transformer network for dual-polarization radar extrapolation. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–15. https://doi.org/10.1109/TGRS.2024.3420417

Zhao, B., Zhang, T., Wang, Y., Lin, P., Liu, H., Huang, P., et al. (2025). Combined LFS and ConvLSTM to forecast marine heatwaves: A case study. *Atmospheric and Oceanic Science Letters*, *19*(2), 100690. https://doi.org/10.1016/j.aosl.2025.100690

Zhao, M., Yang, X.-Q., Tao, L., & Luo, J.-J. (2025). Processes determining the seasonality of accelerated Tibetan Plateau warming during recent decades. *Climate Dynamics*, *63*(2), 116. https://doi.org/10.1007/s00382-025-07596-w

Zhou, L., & Zhang, R. H. (2022). A hybrid neural network model for ENSO prediction in combination with principal oscillation pattern analyses. *Advances in Atmospheric Sciences*, *39*(6), 889–902. https://doi.org/10.1007/s00376-021-1368-4

Zhou, L., & Zhang, R.-H. (2023). A self-attention–based neural network for three-dimensional multivariate modeling and its skillful ENSO predictions. *Science Advances*, *9*(10), eadf2827. https://doi.org/10.1126/sciadv.adf2827

Zhou, X., Zhu, Y., Hou, D., Fu, B., Li, W., Guan, H., et al. (2022). The development of the NCEP global ensemble forecast system version 12. *Weather and Forecasting*, *37*(6), 1069–1084. https://doi.org/10.1175/WAF-D-21-0112.1