TIDE: TENDENCY-INFORMED DYNAMIC ENCODING FOR DETAIL-PRESERVING GLOBAL WEATHER FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Skillful global weather forecasting is essential for agricultural production, energy management, and disaster mitigation. Despite the rapid progress of data-driven forecasting models, two challenges remain insufficiently addressed. First, historical atmospheric states are often compressed through static or weakly conditioned temporal aggregation, limiting the representation of persistent evolution, first-order tendencies, and accelerating changes. Second, repeated spatial downsampling and autoregressive rollout tend to suppress fine-scale structures, resulting in progressively over-smoothed forecasts. To address these challenges, we propose TIDE, a tendency-informed and detail-preserving framework for global weather forecasting. Its State-Tendency-Acceleration Fusion module constructs first- and second-order temporal differences and adaptively integrates them with the latest state across the bottleneck and pathways, therefore better capturing persistent evolution and accelerating atmospheric changes. To preserve spatial structures, TIDE employs a multiresolution analysis and reconstruction architecture that learns temporally evolving detail memories and selectively restores them through gated decoding. Experiments on ERA5 dataset demonstrate that TIDE achieves state-of-the-art performance in global weather forecasting, establishing a new paradigm in this field.

1 INTRODUCTION

Accurate and timely global weather forecasting is essential for agricultural production, energy management, and disaster mitigation. Numerical weather prediction (NWP) has long supported operational forecasting by numerically solving equations governing atmospheric processes (Bauer et al., 2015; Benjamin et al., 2019). However, imperfect representations of complex physical processes can introduce forecast errors (Palmer et al., 2005), while computationally intensive simulations often require substantial supercomputing resources (Bauer et al., 2020).

Meanwhile, high-quality reanalysis datasets, such as ECMWF Reanalysis v5 (ERA5), provide a foundation for learning atmospheric evolution directly from historical states (Hersbach et al., 2020). In recent years, data-driven methods have become an important approach to global weather forecasting owing to their low inference costs and strong predictive capabilities. Existing data-driven weather models explore effective representations of atmospheric dynamics from different perspectives. For example, FourCastNet uses a Fourier Neural Operator to model global spatial relationships (Pathak et al., 2022); Pangu-Weather employs an Earth-specific Transformer for high-resolution three-dimensional weather forecasting (Bi et al., 2023); GraphCast represents global atmospheric states as a multiscale graph and performs message passing using a graph neural network (Lam et al., 2023); and FengWu improves medium-range forecasting through multivariable fusion (Chen et al., 2025). However, they primarily focus on spatial feature extraction and interactions among weather variables, while the temporal evolution of historical atmospheric states and the preservation of details during spatial multiresolution transformations remain insufficiently modeled.

First, historical atmospheric evolution is often represented by static or weakly conditioned temporal aggregation. Existing methods either condition on a single atmospheric state, as in FourCastNet and Pangu-Weather, or fuse two consecutive states through early channel concatenation, as

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

in GraphCast and FengWu, leaving temporal evolution to be learned only implicitly. Even when the input contains a longer historical sequence, such aggregation typically treats information from different time steps uniformly, making it difficult to distinguish the latest atmospheric state, persistent tendencies, and the strengthening or weakening of those tendencies. For three equally spaced historical states, the latest state describes the current weather, the first-order temporal difference captures its primary direction of evolution, and the second-order temporal difference indicates whether this evolution is accelerating, decelerating, or changing direction. If these signals are compressed directly at the input stage, the model may struggle to distinguish steadily persisting circulation systems from rapidly developing local weather processes. Recent studies also highlight the importance of modeling dependencies across multiple temporal scales for global weather forecasting (Tian et al., 2026). This motivates a temporal fusion mechanism that explicitly constructs temporal information of multiple orders and adaptively selects relevant evolution signals based on the current atmospheric state.

Second, fine-scale atmospheric structures are progressively weakened during multiresolution transformations. To control computational complexity on global grids, existing models typically perform repeated spatial downsampling, compressing high-resolution weather fields into a low-resolution bottleneck. However, structures such as frontal boundaries, tropical cyclone cores, jet streams, and strong local gradients are primarily represented by fine-scale features, which can be smoothed or mixed with the large-scale background during repeated downsampling. Similar degradation of spatial details in long-range data-driven forecasting has also been described as model collapse (Li et al., 2026). Existing forecasting models usually reconstruct full-resolution fields through linear patch projection, transposed convolution, or graph-to-grid message passing. Consequently, fine-scale information is carried only implicitly in mixed latent features, rather than maintained as dedicated detail representations throughout spatial compression and reconstruction. To preserve fine-scale structures, a more suitable approach is to let the network learn task-specific multiresolution analysis and reconstruction.

To address these two challenges, we propose TIDE, a Tendency-Informed and DEtail-preserving framework for global weather forecasting. Its core temporal modeling component, State-Tendency-Acceleration (STA) fusion, takes historical atmospheric states at three consecutive time steps and explicitly constructs representations of the latest state, first-order tendency, and second-order acceleration. Instead of simple concatenation or fixed weighting, this module adaptively modulates the contributions of the three temporal components based on the current features, enabling the model to identify persistent evolution and accelerating changes while retaining information about the current weather. This fusion mechanism is applied not only at the lowest-resolution bottleneck but also along multiscale detail pathways, preserving temporal evolution signals at different spatial scales and thereby addressing the first challenge.

To address the second challenge, TIDE further introduces a multiresolution analysis—reconstruction architecture. At each scale, the encoder decomposes features into a semantic representation for large-scale dynamics and a detail memory dedicated to preserving fine-scale information. After several downsampling stages, the bottleneck uses global spatial attention to model large-scale teleconnections. The decoder adopts a hierarchical structure and progressively integrates large-scale features with detail memories. This design follows the analysis—preservation—reconstruction principle of lifting-based multiresolution analysis (Sweldens, 1998) and incorporates shifted-window attention (Liu et al., 2021), allowing the extraction and recovery of fine-scale information to be optimized end to end for weather forecasting.

We evaluate TIDE on ERA5 dataset and compare it with representative data-driven models with consistent temporal coverage and at a spatial resolution of 0.5°. Compared with the best-performing baseline, TIDE achieves higher forecast accuracy across nearly all variables and frequency bands, with peak ACC improvements exceeding 25%. In addition, accurate forecasts of typhoons In-fa and Bavi from different years further demonstrate the importance of enhanced temporal modeling and spatial detail preservation for capturing extreme weather events. In summary, our contributions can be summarized as follows:

- We propose State-Tendency-Acceleration (STA) fusion, which explicitly constructs the latest state, first-order tendency, and second-order acceleration from three historical states and adaptively fuses them at the bottleneck and along detail pathways, improving the representation of persistent evolution and accelerating changes.

- We develop a learnable detail-preserving multiresolution architecture that uses three levels of detail memory and an attention-based gated decoder to analyze, propagate, and reconstruct fine-scale features.
- We conduct comprehensive experiments on global weather forecasting, evaluating TIDE through overall performance comparisons, frequency-band analysis, extreme weather case studies, and ablation studies to validate the effectiveness of temporal modeling with multiple orders and spatial detail preservation.

2 PRELIMINARIES

2.1 PROBLEM STATEMENT

Global weather forecasting aims to predict future atmospheric states from historical observations. Let $X_t \in \mathbb{R}^{C\times H\times W}$ denote the weather state at time t, where C is the total number of channels comprising surface variables and upper-air variables at different pressure levels, and H and W are the numbers of latitude and longitude grid points, respectively. Given three historical states sampled at a fixed interval δ, we define

$$\mathcal{H}_t = (X_{t-2\delta}, X_{t-\delta}, X_t)\,. \tag{1}$$

Our objective is to learn a parameterized model f_θ that predicts the weather state at the next time step:

$$\widehat{X}_{t+\delta} = f_\theta\left(\mathcal{H}_t\right). \tag{2}$$

For a forecast lead time of $K\delta$, we apply f_θ autoregressively, updating the three-state window after each step by appending the prediction and removing the earliest state, until $\widehat{X}_{t+K\delta}$ is obtained.

2.2 RELATED WORK

Data-driven Global Weather Forecasting. In recent years, data-driven models have substantially advanced global medium-range weather forecasting by learning atmospheric evolution directly from reanalysis data. FourCastNet uses an Adaptive Fourier Neural Operator to model global spatial dependencies in the frequency domain (Pathak et al., 2022); Pangu-Weather represents three-dimensional atmospheric states using an Earth-specific Transformer and employs models with different prediction intervals to reduce error accumulation at longer lead times (Bi et al., 2023); and GraphCast maps a regular latitude-longitude grid onto a multiscale spherical mesh, capturing long-range atmospheric interactions through graph message passing (Lam et al., 2023). In addition, FengWu uses multimodal encoding and fusion to model interactions among meteorological variables (Chen et al., 2025). These studies establish effective frameworks for global weather modeling through frequency-domain operations, three-dimensional attention, graph representations, and multimodal learning. Their main focus, however, remains spatial feature extraction, interactions among weather variables, and strategies for forecasting at longer lead times.

Multiscale Temporal Modeling. Atmospheric evolution differs considerably across temporal scales, motivating recent work to strengthen temporal modeling through prediction intervals and long-term background states. ARROW learns shared and interval-specific dynamics across multiple forecast intervals within a unified Multi-Interval Forecasting Model and uses an Adaptive Rollout Scheduler to dynamically select autoregressive step sizes based on the current weather state (Tian et al., 2026). TianQuan-S2S targets subseasonal-to-seasonal forecasting and combines five days of historical weather states with a multiyear climatological mean to mitigate the diminishing influence of initial conditions as lead time increases (Li et al., 2026). These methods focus, respectively, on model sharing across prediction intervals and rollout decisions, and on climatological background information for long-range forecasting. Complementing these approaches, TIDE investigates temporal relationships of multiple orders among consecutive historical states for a given prediction interval. By explicitly constructing state, tendency, and acceleration and adaptively fusing them at multiple spatial resolutions, TIDE represents the current atmospheric pattern, the direction of persistent change, and the strengthening or weakening of that change.

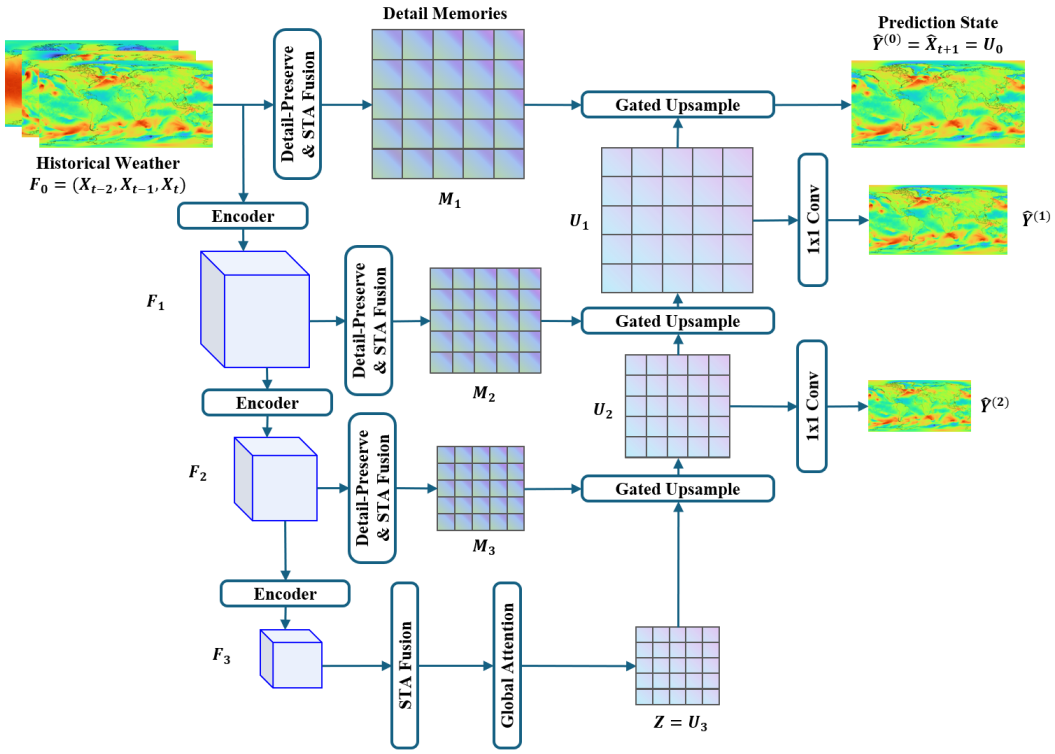

Figure 1: Overview of the TIDE framework. Historical weather states at three consecutive time steps undergo multiple encoding stages to progressively extract main-path features and detail components. Temporal modeling and spatial attention are then applied at the bottleneck, followed by progressive gated reconstruction using the retained detail memories to predict the future weather field.

3 METHODOLOGY

We propose TIDE, a Tendency-Informed and DEtail-preserving framework for global weather forecasting. Given three historical weather fields in chronological order, the model explicitly constructs state, tendency, and acceleration features and generates the next weather field through hierarchical spatial encoding, detail memory fusion, and progressive reconstruction. Figure 1 illustrates the overall architecture.

3.1 MODEL OVERVIEW

Given three historical weather fields ordered from earliest to latest, the model input is denoted by $\mathcal{X} \in \mathbb{R}^{B\times 3\times C\times H\times W}$, where B is the batch size, C is the number of variable channels, and H and W are the latitude and longitude grid dimensions, respectively. The model learns a mapping from the historical sequence to the weather field at the next time step, $f_\theta : \mathcal{X} \mapsto \widehat{Y}$, where $\widehat{Y} \in \mathbb{R}^{B\times C\times H\times W}$.

To summarize the overall information flow, let $F^0 = \mathcal{X}$, and let F^ℓ denote the features at encoding scale ℓ, which retain the temporal dimension of three frames. The three-level encoder can be expressed as

$$F^\ell = \mathcal{E}_\ell\big(F^{\ell-1}\big), \qquad \ell = 1, \ldots, L, \quad L = 3, \tag{3}$$

where $\mathcal{E}_\ell$ denotes the spatial downsampling and feature refinement operations at level ℓ, with parameters shared across the three historical frames. The model also constructs a detail memory M^ℓ from $F^{\ell-1}$ at the corresponding scale. Denoting the detail extraction, STA fusion, global attention, and

gated reconstruction modules by $\mathcal{D}_\ell$, $\mathcal{S}_\ell$, $\mathcal{A}$, and $\mathcal{U}$, respectively, the overall prediction process is

$$\begin{aligned} M^\ell &= \mathcal{S}_\ell\big(\mathcal{D}_\ell(F^{\ell-1})\big), \qquad \ell = 1, \dots, L, \\ Z &= \mathcal{A}\big(\mathcal{S}_{\mathrm{b}}(F^L)\big), \\ \widehat{Y} &= \mathcal{U}\big(Z, \{M^\ell\}_{\ell=1}^{L}\big). \end{aligned} \tag{4}$$

Here, $\mathcal{S}_{\mathrm{b}}$ is a separate STA fusion module at the bottleneck, and $\mathcal{U}$ denotes coarse-to-fine upsampling, gated detail injection, and final field reconstruction. The internal computations of these mappings are described in the following subsections.

During training, in addition to the full-resolution prediction $\widehat{Y}^{(0)} = \widehat{Y}$, the decoder produces auxiliary predictions at half and quarter resolution, denoted by $\widehat{Y}^{(1)}$ and $\widehat{Y}^{(2)}$. Let $Y^{(0)} = Y$; the corresponding supervision targets $Y^{(1)}$ and $Y^{(2)}$ are constructed successively using area interpolation. For odd grid dimensions, the actual spatial dimensions at scale s are $H_s = \lceil H/2^s \rceil$ and $W_s = \lceil W/2^s \rceil$, where $s \in \{0, 1, 2\}$. Following the use of auxiliary supervision at intermediate layers (Lee et al., 2015), the total loss is defined as

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{lat}}\left(\widehat{Y}^{(0)}, Y^{(0)}\right) + r_1\, \mathcal{L}_{\mathrm{lat}}\left(\widehat{Y}^{(1)}, Y^{(1)}\right) + r_2\, \mathcal{L}_{\mathrm{lat}}\left(\widehat{Y}^{(2)}, Y^{(2)}\right), \tag{5}$$

where $\mathcal{L}_{\mathrm{lat}}$ is the cosine-latitude-weighted mean squared error, with $r_1 = 0.25$ and $r_2 = 0.1$. Specifically, for a prediction grid of height H', the weight of row h is defined as

$$w_h = \frac{\cos(\phi_h)}{\frac{1}{H'}\sum_{r=1}^{H'} \cos(\phi_r)}, \qquad \mathcal{L}_{\mathrm{lat}}(\widehat{Y}, Y) = \frac{1}{BCH'W'} \sum_{b,c,h,w} w_h \big(\widehat{Y}_{b,c,h,w} - Y_{b,c,h,w}\big)^2, \tag{6}$$

where ϕ_h is the latitude corresponding to row h.

3.2 State-Tendency-Acceleration Fusion

To explicitly characterize temporal changes in feature space, the State-Tendency-Acceleration (STA) fusion module $\mathcal{S}$ takes a feature sequence $Q = (q_0, q_1, q_2)$ at the same scale, ordered from earliest to latest, and constructs three components: state, tendency, and acceleration:

$$S(Q) = q_2, \qquad T(Q) = q_2 - q_1, \qquad A(Q) = q_2 - 2q_1 + q_0. \tag{7}$$

The state component retains the features at the latest time step, the tendency component is the first-order difference between consecutive frames, and the acceleration component is the second-order difference across three frames.

Let $q_i \in \mathbb{R}^{B \times C_q \times H_q \times W_q}$. At each grid location, the STA fusion module concatenates the three components along the channel dimension and fuses them using a multilayer perceptron (MLP) with shared parameters:

$$\mathcal{S}(Q) = \mathrm{Linear}_{2C_q \to C_q}\left(\mathrm{GELU}\left[\mathrm{Linear}_{3C_q \to 2C_q}\big([S(Q); T(Q); A(Q)]\big)\right]\right). \tag{8}$$

Thus, STA fusion maps temporal features with $3C_q$ channels back to C_q channels while preserving the spatial dimensions of the feature map. The mapping parameters are shared across all grid locations within the same scale, while the detail memories at different scales and the bottleneck representation use separate fusion parameters.

3.3 Learnable Detail-preserving Multiresolution Architecture

The hierarchical encoder $\mathcal{E}_\ell$ consists of Patch Merging and window-attention refinement. Patch Merging combines adjacent 2×2 feature patches and then adjusts the channel dimension using LayerNorm and a learnable linear projection. When a spatial dimension is odd, boundary padding is applied before merging. Let d be the number of bottleneck channels; the channel dimensions at the three encoding levels are $C_1 = d/4$, $C_2 = d/2$, and $C_3 = d$, respectively. After each merging operation, the model successively applies regular-window and shifted-window attention blocks with learnable relative position biases. This local feature refinement follows the Swin Transformer (Liu et al., 2021). The encoder parameters at each level are shared across the three historical frames.

The detail extraction module $\mathcal{D}_\ell$ takes the three-frame features from the preceding encoding scale and applies the same learnable downsampling operation to each frame. It first extracts local features using a 3×3 depthwise convolution followed by GELU, then reduces spatial resolution using a 2×2 depthwise convolution with a stride of 2, and finally mixes channel information using a 1×1 convolution followed by GELU. The output has the same spatial dimensions as the main-path features at level ℓ. Subsequently, $\mathcal{S}_\ell$ fuses the three-frame detail features into a single M^ℓ, which serves as the detail memory for the corresponding decoding stage.

At the deepest level, the bottleneck first applies STA fusion to F^L and adds learnable spatial positional embeddings. The resulting features are flattened into a sequence of spatial tokens and processed by global attention blocks consisting of multihead self-attention and position-wise MLPs before being reshaped into the two-dimensional feature map Z. Each attention block uses pre-LayerNorm and residual connections, with global interactions following the basic Transformer architecture (Vaswani et al., 2017). The bottleneck models cross-regional relationships at a lower spatial resolution and passes the resulting representation to the progressive decoder.

The decoder $\mathcal{U}$ starts from Z and uses the detail memories M^ℓ in the order $\ell = L, L-1, \ldots, 1$. Let $U^L = Z$. At decoding level ℓ, the current features are first projected to the same channel dimension as M^ℓ, and gating weights are computed from both representations. The main-path features and gated detail memory are then upsampled separately and added at the target scale:

$$\begin{aligned} G^\ell &= \mathrm{Sigmoid}\Big(\mathrm{GateMap}^\ell\big([\mathrm{ChannelProj}^\ell(U^\ell); M^\ell]\big)\Big), \\ U^{\ell-1} &= \mathrm{Refine}^\ell\Big(\mathrm{Upsample}^\ell_{\mathrm{main}}(U^\ell) + \mathrm{Upsample}^\ell_{\mathrm{detail}}(G^\ell \odot M^\ell)\Big). \end{aligned} \tag{9}$$

Here, $[\,;\,]$ denotes channel-wise concatenation, and $\odot$ denotes element-wise multiplication. The gating weights G^ℓ select the detail information to inject based on the current decoder features. After upsampling, the main pathway and detail pathway are added and refined to produce the decoder features $U^{\ell-1}$ at the next scale. The first two decoding stages use PixelShuffle-based channel-to-space rearrangement (Shi et al., 2016), followed by window-attention refinement. The final stage uses bilinear interpolation and convolution to restore full resolution. The final prediction is $\widehat{Y}^{(0)} = U^0$, while auxiliary prediction heads applied to the intermediate features U^1 and U^2 produce $\widehat{Y}^{(1)}$ and $\widehat{Y}^{(2)}$, respectively, for the multiscale auxiliary training objective.

4 Experiments

4.1 Experimental Settings

We construct the global weather forecasting dataset using ECMWF Reanalysis v5 (ERA5) (Hersbach et al., 2020), downloaded from WeatherBench 2 (Rasp et al., 2024). Owing to limited GPU resources, we downsample the original 0.25° latitude-longitude grid to 0.5°, yielding a global grid with $H = 361$ and $W = 720$ for training and inference. Each grid point contains 11 meteorological variables: 10 m zonal wind $U10$, 10 m meridional wind $V10$, 2 m temperature $T2M$, mean sea level pressure MSL, and total precipitation TP, together with geopotential Z, specific humidity Q, relative humidity R, temperature T, zonal wind U, and meridional wind V at 13 pressure levels. This gives a total of $C = 5 + 6 \times 13 = 83$ channels. The pressure levels are 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, and 1000 hPa. The data are sampled at 6 h intervals, and the model uses three consecutive historical weather fields (X_{t-2}, X_{t-1}, X_t) to predict X_{t+1}. The training, validation, and test sets cover 1990–2018, 2019–2020, and 2021, respectively, ensuring temporal separation.

The comparison methods include Pangu-Weather, FengWu, and GraphCast (Bi et al., 2023; Chen et al., 2025; Lam et al., 2023). All three baselines are trained from scratch on the same dataset, with their core network architectures kept as close as possible to the original designs. To achieve the best forecast accuracy, Pangu-Weather uses a 24 h prediction interval, while the other models use 6 h intervals. Because Pangu-Weather and GraphCast match or exceed the operational deterministic numerical weather prediction systems of their time on several benchmarks reported in their respective papers (Bi et al., 2023; Lam et al., 2023), our main comparisons focus on data-driven models under the same data conditions and do not additionally include numerical weather prediction systems.

Following prior work (Bi et al., 2023), we use latitude-weighted root mean square error (RMSE) and anomaly correlation coefficient (ACC) as the primary evaluation metrics. To distinguish forecast skill at different spatial scales, we further define frequency-band RMSE. Using the spherical harmonic transform $\mathcal{S}$ on an equiangular latitude-longitude grid, we retain coefficients within a specified frequency band according to the total spherical harmonic wavenumber ℓ and reconstruct the corresponding spatial field using the inverse transform (Bonev et al., 2023). Let M_k be the mask for frequency band k; the band operator is defined as

$$\mathcal{B}_k(f) = \mathcal{S}^{-1}(M_k \odot \mathcal{S}(f)) \,. \tag{10}$$

We use five nonoverlapping bands, L1–L5, covering ℓ = 0–20, 21–80, 81–200, 201–280, and 281–360, respectively, in increasing order of spatial frequency. Let $E_k^{(n)} = \mathcal{B}_k(\widehat{y}^{(n)} - y^{(n)})$. Then,

$$\mathrm{RMSE}_k = \left[\frac{1}{NHW} \sum_{n,i,j} w_i \big(E_{k,ij}^{(n)}\big)^2 \right]^{1/2} . \tag{11}$$

Low-frequency bands primarily reflect forecasts of large-scale circulation and global weather tendencies, whereas high-frequency bands emphasize smaller-scale gradients and local structures. Therefore, lower frequency-band RMSE in L4–L5 demonstrates that the model preserves useful spatial details.

4.2 Global Forecasting

Figure 2 summarizes global forecasting results on the test set, reporting RMSE and ACC for representative surface and upper-air variables at different lead times. Lower RMSE and higher ACC indicate more accurate forecasts. For the 7-day forecast of Z500, TIDE reduces RMSE by more than 20% and improves ACC by more than 25% compared with the best-performing baseline. TIDE achieves higher forecast accuracy and lower errors for nearly all weather variables.

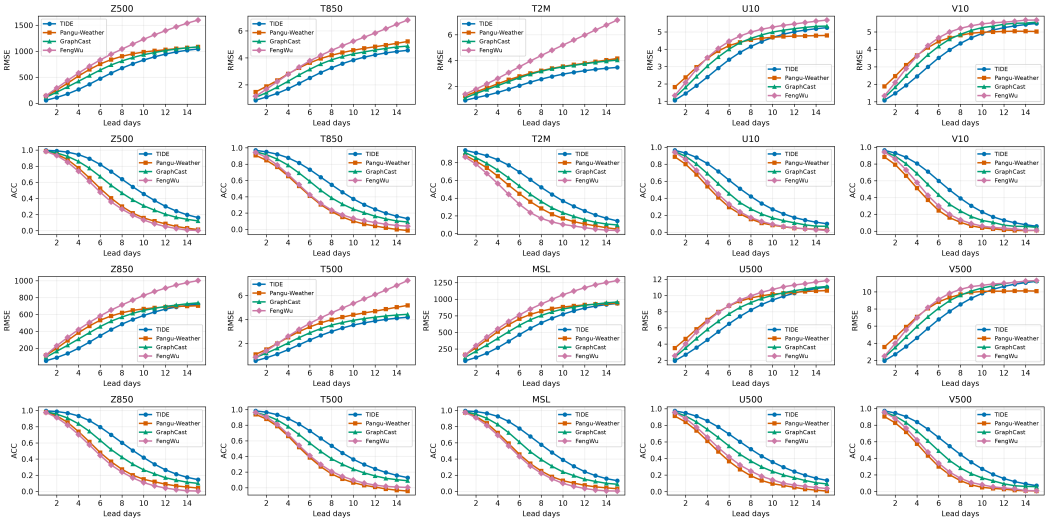

Figure 2: Comparison of overall forecast performance. Following Pangu-Weather (Bi et al., 2023), we use the same representative weather variables to compare forecast accuracy across models.

Figure 3 presents the frequency-band RMSE results. Lower total spherical harmonic wavenumbers mainly describe planetary-scale circulation and broad weather patterns, whereas higher wavenumbers emphasize fronts, gradients around cyclones, and more localized field structures. Except for some variables in the low-frequency bands L1–L2, TIDE achieves better forecast quality across most variables and frequency bands. In particular, for high-frequency details in L4–L5, TIDE reduces Z500 RMSE by up to 50% compared with the best-performing baseline, demonstrating that detail memory and multiresolution reconstruction can reliably improve forecasts of smaller-scale structures.

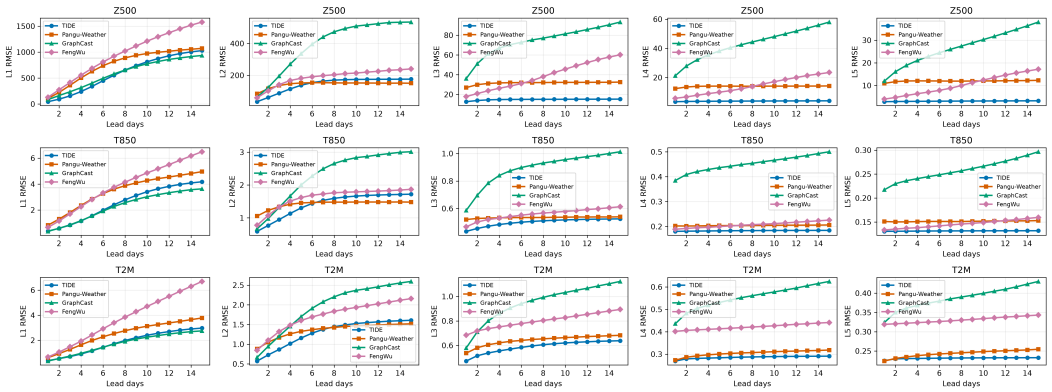

Figure 3: Frequency-band RMSE for selected weather variables. Lower RMSE in bands L4–L5 indicates better modeling of spatial details.

4.3 Case Studies

Typhoons are representative weather systems characterized by rapid changes in atmospheric conditions within local regions. Their low-pressure centers, strong wind circulations, and precipitation distributions exhibit pronounced spatial gradients and local structures. Typhoon case studies therefore allow us to assess a model's ability to represent and preserve local spatial details through its track forecasting performance. We select Typhoon In-fa in 2021 and Typhoon Bavi in 2026 as case studies. Reference typhoon tracks are obtained from data released by the Japan Meteorological Agency (JMA), which are more accurate than tracks manually calculated from ERA5.

In-fa, the sixth typhoon of 2021, made landfall in Pinghu, Zhejiang, China, on July 26, affecting 4.82 million people and causing direct economic losses of RMB 13.2 billion (Ministry of Emergency Management of the People's Republic of China, 2022). Figure 4 compares the initial field, observed track, and forecast tracks from different models for Typhoon In-fa in 2021, together with track errors as a function of lead time. TIDE successfully predicts landfall in the Jiangsu-Zhejiang region four days in advance. In comparison, GraphCast incorrectly estimates the landfall time, while the tracks predicted by the other baselines diverge completely from the observed track.

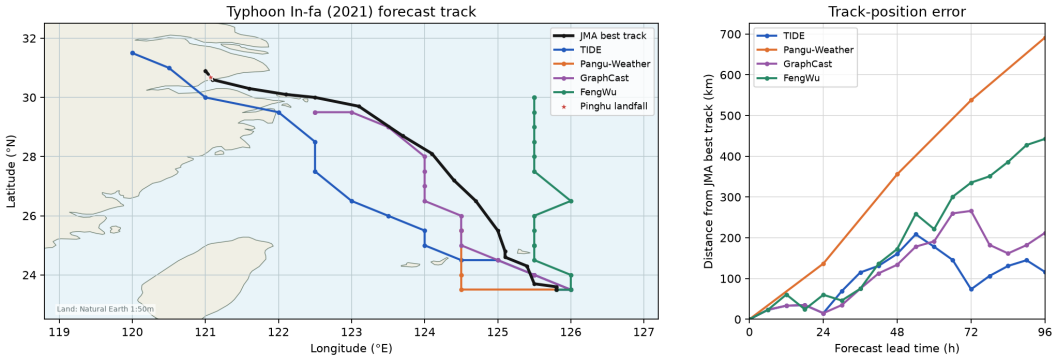

Figure 4: Visualization of forecast tracks and track errors for Typhoon In-fa.

Bavi, the ninth typhoon of 2026, made landfall along the coast of Yuhuan, Zhejiang, China, on the evening of July 11 (Taihu Basin Authority, Ministry of Water Resources of China, 2026), becoming the strongest typhoon to make landfall in Zhejiang in July since 1949 (Ministry of Emergency Management of the People's Republic of China, 2026). Figure 5 examines the case of Typhoon Bavi in 2026. TIDE successfully predicts landfall along the Zhejiang coast. In comparison, GraphCast incorrectly predicts landfall in Taiwan, Pangu-Weather incorrectly estimates the landfall time, and the tracks predicted by the other baselines deviate substantially from the observed track.

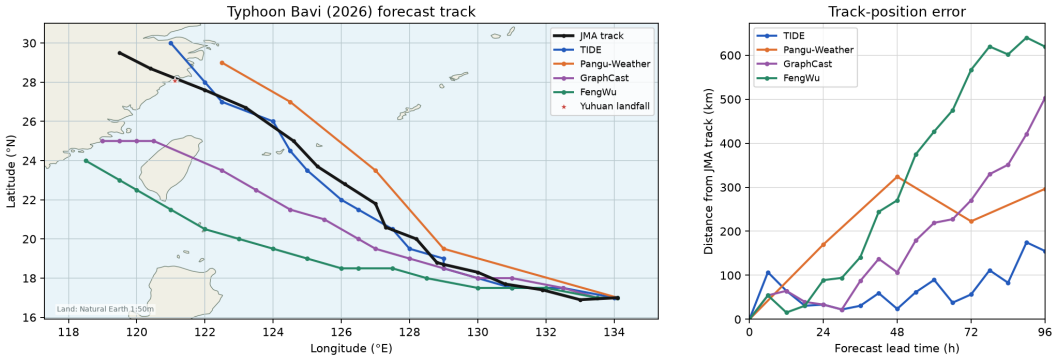

Figure 5: Visualization of forecast tracks and track errors for Typhoon Bavi.

4.4 Ablation Studies

Effect of Temporal Aggregation. To assess the role of temporal modeling, we keep the encoder, detail branch, and decoder unchanged and separately train variants using only the latest state X_t, two consecutive frames (X_{t-1}, X_t), and all three frames (X_{t-2}, X_{t-1}, X_t). To assess the role of explicit differences, we also include a control variant that directly concatenates the three raw frames. Table 1 shows that explicit differences make an important contribution to accurate weather state prediction.

Table 1: Effects of STA fusion, detail preservation, and the multiscale loss. All values are evaluated at a 72-hour lead time.

Methods	Z500		T850		T2M	
	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑
Full model	**180.81**	**0.97**	**1.37**	**0.92**	**1.30**	**0.88**
Latest state only (X_t)	287.11	0.93	1.70	0.87	1.72	0.81
Two frames (X_{t-1}, X_t)	241.37	0.95	1.49	0.90	1.41	0.85
Raw three-frame concatenation	201.91	0.96	1.39	0.92	1.33	0.87
Without detail memory	244.19	0.95	1.56	0.89	1.42	0.85
Direct addition without gating	198.77	0.96	1.42	0.91	1.33	0.87
Full-resolution loss only	260.09	0.94	1.46	0.90	1.39	0.85

Effect of Detail Preservation and Reconstruction. Table 1 further compares two variants related to detail reconstruction: removing detail memory and replacing gated injection with direct addition. The results demonstrate the importance of accurately preserving spatial details for forecast performance.

Effect of the Multiscale Loss. We compare a variant trained using only the full-resolution loss with the model trained using supervision at all three scales. Table 1 shows that multiscale training signals make an important contribution to forecast performance.

5 Conclusion

We introduced TIDE, a multiresolution global weather forecasting framework combining State-Tendency-Acceleration (STA) fusion with learnable detail memory and gated reconstruction. Experiments on ERA5 at 0.5° resolution demonstrate state-of-the-art forecasting performance, supported by frequency-band evaluations, typhoon case studies, and ablation analyses. Future work will investigate error accumulation in long-range autoregressive forecasting and explore the integration of physical constraints and additional Earth system information to improve extreme weather forecasting.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation. We did not use these tools to develop theoretical models or conceptual frameworks, design or provide feedback on the research methodology or experiments, implement methods, clean or reformat datasets, support qualitative or thematic data analysis, or interpret results. The following activities are not applicable to this work: generating synthetic datasets, formulating mathematical claims, providing essential steps in mathematical proofs, assisting with the writing of proofs, and proposing or refining hypotheses. We have reviewed all AI-assisted material and take full responsibility for the final content of this work, including any text, claims, or artifacts produced with the assistance of generative AI.

REPRODUCIBILITY STATEMENT

To ensure reproducibility, we use the publicly available and widely adopted ERA5 dataset to train, validate, and test TIDE and all baselines. The Methodology section describes the model's temporal modeling and spatial detail preservation and reconstruction mechanisms in detail, including their mathematical formulations and implementation as deep neural networks. We plan to release the training code on GitHub upon acceptance to further promote the integration and application of machine learning in operational weather forecasting.

REFERENCES

Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015. doi: 10.1038/nature14956. URL https://doi.org/10.1038/nature14956.

Peter Bauer, Tiago Quintino, Nils Wedi, Antonio Bonanni, Marcin Chrust, Willem Deconinck, Michail Diamantakis, Peter Düben, Stephen English, Johannes Flemming, et al. *The ECMWF scalability programme: Progress and plans*. European Centre for Medium Range Weather Forecasts, 2020.

Stanley G Benjamin, John M Brown, Gilbert Brunet, Peter Lynch, Kazuo Saito, and Thomas W Schlatter. 100 years of progress in forecasting and nwp applications. *Meteorological Monographs*, 59:13–1, 2019.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619:533–538, 2023. doi: 10.1038/s41586-023-06185-3.

Boris Bonev, Thorsten Kurth, Christian Hundt, Jaideep Pathak, Maximilian Baust, Karthik Kashinath, and Anima Anandkumar. Spherical fourier neural operators: Learning stable dynamics on the sphere. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 2806–2823. PMLR, 2023. URL https://proceedings.mlr.press/v202/bonev23a.html.

Kang Chen, Tao Han, Fenghua Ling, Junchao Gong, Lei Bai, Xinyu Wang, Jing-Jia Luo, Ben Fei, Wenlong Zhang, Xi Chen, Leiming Ma, Tianning Zhang, Rui Su, Yuanzheng Ci, Bin Li, Xiaokang Yang, and Wanli Ouyang. The operational medium-range deterministic weather forecasting can be extended beyond a 10-day lead time. *Communications Earth & Environment*, 6:518, 2025. doi: 10.1038/s43247-025-02502-y. URL https://doi.org/10.1038/s43247-025-02502-y.

Hans Hersbach, Bill Bell, Paul Berrisford, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi: 10.1002/qj.3803.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, et al. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023. doi: 10.1126/science.adi2336.

Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pp. 562–570. PMLR, 2015. URL https://proceedings.mlr.press/v38/lee15a.html.

Guowen Li, Xintong Liu, Yang Liu, Mengxuan Chen, Shilei Cao, Xuehe Wang, Juepeng Zheng, Jinxiao Zhang, Haoyuan Liang, Lixian Zhang, Jiuke Wang, Meng Jin, Hong Cheng, and Haohuan Fu. TianQuan-S2S: A subseasonal-to-seasonal global weather model via incorporate climatology state. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ad15848baa3932c0d2deabf0e11d1dcd-Abstract-Conference.html.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10012–10022, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Liu_Swin_Transformer_Hierarchical_Vision_Transformer_Using_Shifted_Windows_ICCV_2021_paper.html.

Ministry of Emergency Management of the People's Republic of China. The ministry of emergency management releases the top ten natural disasters of 2021, January 2022. URL https://www.mem.gov.cn/xw/yjglbgzdt/202201/t20220123_407199.shtml. Published January 23, 2022. In Chinese.

Ministry of Emergency Management of the People's Republic of China. National natural disaster situation in july 2026, August 2026. URL https://www.mem.gov.cn/xw/yjglbgzdt/202608/t20260814_680668.shtml. Published August 14, 2026. In Chinese; title translated into English.

TN Palmer, GJ Shutts, R Hagedorn, FJ Doblas-Reyes, Thomas Jung, and M Leutbecher. Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.*, 33(1): 163–193, 2005.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar. FourCastNet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022. doi: 10.48550/arXiv.2202.11214. URL https://arxiv.org/abs/2202.11214.

Stephan Rasp, Stephan Hoyer, Alexander Merose, Ian Langmore, Peter Battaglia, Tyler Russell, Alvaro Sanchez-Gonzalez, Vivian Yang, Rob Carver, Shreya Agrawal, Matthew Chantry, Zied Ben Bouallègue, Peter Düben, Carla Bromberg, Jared Sisk, Luke Barrington, Aaron Bell, and Fei Sha. WeatherBench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems*, 16:e2023MS004019, 2024. doi: 10.1029/2023MS004019. URL https://doi.org/10.1029/2023MS004019.

Wenzhe Shi, Jose Caballero, Ferenc Huszar, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1874–1883, 2016. URL https://openaccess.thecvf.com/content_cvpr_2016/html/Shi_Real-Time_Single_Image_CVPR_2016_paper.html.

Wim Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM Journal on Mathematical Analysis*, 29(2):511–546, 1998. doi: 10.1137/S0036141095289051. URL https://doi.org/10.1137/S0036141095289051.

Taihu Basin Authority, Ministry of Water Resources of China. Latest developments of typhoon bavi and outlook for rainfall and water conditions (iii), July 2026. URL https://www.tba.gov.cn/slbthlyglj/thyw/content/c30cc5a2-4d3f-4491-aab1-22fe22d115fd.html. Published July 12, 2026. In Chinese; title translated into English.

Jindong Tian, Yifei Ding, Ronghui Xu, Hao Miao, Chenjuan Guo, and Bin Yang. ARROW: An adaptive rollout and routing method for global weather forecasting. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/59fb5cf4683eaec4c56918a69637c9d2-Abstract-Conference.html.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

A NOTATION

Table 2: Summary of notation used in this paper.

Symbol	Description
Model and training	
$\mathcal{X}$	Batch of three historical weather fields ordered from earliest to latest.
X_t	Weather field at time step t; X_{t-2}, X_{t-1}, X_t form the input sequence, and X_{t+1} is the prediction target.
B	Batch size.
C	Number of weather-variable channels; $C = 83$ in the experiments.
H, W	Numbers of latitude and longitude grid points, respectively.
f_θ, θ	One-step forecasting model and its trainable parameters.
$Y, \widehat{Y}$	Ground-truth next weather field and its prediction.
L	Number of encoding scales; $L = 3$.
ℓ	Encoder or decoder scale index, $\ell \in \{1, \dots, L\}$.
H_ℓ, W_ℓ	Spatial dimensions at encoding scale ℓ, where $H_\ell = \lceil H/2^\ell \rceil$ and $W_\ell = \lceil W/2^\ell \rceil$.
d	Bottleneck channel width; the three encoder widths are $C_1 = d/4$, $C_2 = d/2$, and $C_3 = d$.
F^0, F^ℓ	Input sequence $F^0 = \mathcal{X}$ and the three-frame feature sequence after encoder scale ℓ.
$\mathcal{E}_\ell$	Per-frame encoder at scale ℓ, consisting of Patch Merging and Swin refinement.
$\mathcal{D}_\ell$	Learnable detail-downsampling operation applied to $F^{\ell-1}$.
$Q = (q_0, q_1, q_2)$	Three feature maps supplied to an STA module in chronological order.
C_q, H_q, W_q	Channel count and spatial dimensions of each feature map q_i.
$S(Q)$	State component of STA fusion, defined as q_2.
$T(Q)$	Tendency component of STA fusion, defined as $q_2 - q_1$.
$A(Q)$	Acceleration component of STA fusion, defined as $q_2 - 2q_1 + q_0$.
$\mathcal{S}_\ell$	STA fusion module for the detail branch at scale ℓ.
$\mathcal{S}_\mathrm{b}$	Independent STA fusion module at the bottleneck.
M^ℓ	STA-fused detail memory at scale ℓ.
$\mathcal{A}$	Global-attention operation at the bottleneck.
Z	Bottleneck representation after STA fusion and global attention; $U^L = Z$.
$\mathcal{U}$	Overall gated decoder mapping Z and $\{M^\ell\}_{\ell=1}^L$ to $\widehat{Y}$.
U^ℓ	Decoder feature at scale ℓ; $U^0 = \widehat{Y}$.
G^ℓ	Sigmoid gate that controls the contribution of M^ℓ to decoder stage ℓ.
$\mathrm{ChannelProj}^\ell$	Channel projection of U^ℓ before gate computation.
$\mathrm{GateMap}^\ell$	Mapping from the concatenated decoder context and detail memory to gate logits.
$\mathrm{Upsample}^\ell_{\mathrm{main}}$	Upsampling operation for the main decoder path.
$\mathrm{Upsample}^\ell_{\mathrm{detail}}$	Upsampling operation for the gated detail path.
Refine^ℓ	Feature refinement after the two upsampled paths are combined.
$[\,;\,], \odot$	Channel-wise concatenation and element-wise multiplication, respectively.

Table 2 (continued)

Symbol	Description
s	Supervision-scale index: $s = 0, 1, 2$ denotes full, half, and quarter resolution.
H_s, W_s	Spatial dimensions of supervision scale s.
$\widehat{Y}^{(s)}, Y^{(s)}$	Prediction and corresponding ground-truth target at supervision scale s.
$\mathcal{L}_{\mathrm{lat}}$	Cosine-latitude-weighted mean squared error.
$\mathcal{L}_{\mathrm{total}}$	Weighted sum of the full-resolution and two auxiliary training losses.
Evaluation	
n, N	Evaluation-sample index and total number of evaluation samples.
i, j	Latitude-row and longitude-column indices.
ϕ_i	Latitude of grid row i.
w_i	Cosine-latitude weight for row i, normalized to have a mean of one.
$\widehat{y}_{ij}^{(n)}, y_{ij}^{(n)}$	Predicted and true values for a given variable, lead time, sample, and grid point.
$\mu_{ij}^{(n)}$	Climatological value matched to sample n and grid point (i, j).
$\widehat{a}_{ij}^{(n)}, a_{ij}^{(n)}$	Predicted and true anomalies after subtracting the same climatology.
RMSE, ACC	Latitude-weighted root mean square error and anomaly correlation coefficient.
$\mathcal{S}, \mathcal{S}^{-1}$	Spherical harmonic transform on the equiangular grid and its inverse.
k	Frequency-band index, $k \in \{1, \ldots, 5\}$.
L1–L5	Five bands of increasing spatial frequency, covering spherical harmonic degrees 0–20, 21–80, 81–200, 201–280, and 281–360, respectively.
M_k	Spectral mask that retains spherical harmonic coefficients in band k.
$\mathcal{B}_k$	Band-pass operator formed by a spherical harmonic transform, multiplication by M_k, and an inverse transform.
$E_k^{(n)}$	Band-filtered prediction-error field for sample n.
RMSE_k	Latitude-weighted RMSE evaluated in frequency band k.
Weather-variable abbreviations	
U10, V10	Zonal and meridional wind at 10 m, respectively.
T2M	Air temperature at 2 m.
MSL	Mean sea-level pressure.
TP	Total precipitation for the specified accumulation period.
Z, Q, R, T, U, V	Geopotential, specific humidity, relative humidity, air temperature, zonal wind, and meridional wind at the selected pressure levels, respectively.