PMGC-SimVP: Parametric Multi-scale Gated Convolution for Global Ionospheric TEC Prediction

Yu Xia[1,2][0009-0005-1402-0429], Yingkui Gong[1,2⋆], and Hao Zhang[1,2]

[1] Chinese Academy of Sciences, Aerospace Information Research Institute, Beijing, China

[2] School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China

Abstract. Spatiotemporal prediction of complex dynamical systems depends not only on historical observations but also on evolving external control parameters. Crucially, for ionospheric Total Electron Content, existing concatenation-based conditioning fails to capture region-dependent responses driven by space weather.

We propose PMGC-SimVP, a spatiotemporal prediction framework with *Parametric Multi-scale Gated Convolution*. PMGC injects external control parameters into local dynamics via latitudinally structured modulation, enabling explicit modeling of non-stationary evolution and regional heterogeneity. Built upon a purely convolutional backbone and a low-rank temporal adapter, PMGC-SimVP remains computationally efficient.

We evaluate PMGC-SimVP on global TEC forecasting, where dynamics are strongly driven by space weather and exhibit pronounced latitudinal variability. The proposed method achieves consistent improvements over recent state-of-the-art models, including Transformer-based and frequency-domain approaches, across overall evaluation settings and during geomagnetic disturbances. Experiments on generic video prediction benchmarks further verify the robustness of the proposed backbone even without parameter injection.

Keywords: Spatiotemporal Prediction · Ionospheric TEC · Parameter Conditioning

1 Introduction

The ionosphere is a highly dynamic plasma system continuously driven by solar radiation and geomagnetic activity. Its state is commonly characterized by the global Total Electron Content (TEC), whose spatiotemporal variability modulates radio wave propagation, causing GNSS signal delays and positioning errors and degrading high-frequency and satellite communications [2, 15, 32]. Accurate

⋆ Corresponding author.

and stable TEC forecasting is therefore critical for a range of practical applications, including GNSS error correction and integrity monitoring, real-time space weather monitoring and early warning, and the reliability of satellite-based navigation and communication systems [4, 28]. Yet achieving reliable long-horizon forecasting at the global scale remains highly challenging due to strong non-stationarity and complex coupling mechanisms in the ionosphere [2, 15].

Deep learning has achieved notable progress in spatiotemporal prediction, ranging from recurrent and memory-augmented models to efficient convolutional predictors and Transformer/frequency-domain designs [8–10, 22, 36, 38]. Data-driven TEC forecasting has also been explored and shows promising short-term performance across different architectures [18, 19, 26, 27]. Nevertheless, directly transferring generic predictors to global TEC often remains suboptimal.

Global TEC forecasting poses several intrinsic challenges. First, enhanced solar activity and geomagnetic storms induce strong non-stationarity, which can severely degrade history-based extrapolation [2, 19]. Second, pronounced latitudinal heterogeneity yields distinct regional responses under identical forcing, while representing global geophysical fields on latitude–longitude grids introduces boundary discontinuities and scale distortions [6, 7, 13]. Finally, long-horizon prediction is prone to error accumulation and spatial over-smoothing, undermining forecast stability [38].

These characteristics motivate modeling global TEC forecasting as a *conditional spatiotemporal prediction* problem, where evolution depends on historical states and time-varying external drivers [2,15,32]. While conditional modulation has been explored in generic vision tasks [5, 23, 33, 39], efficient and stable integration of physical drivers into spatiotemporal predictors remains underexplored for global geophysical fields.

We propose PMGC-SimVP (Parametric Multi-scale Gated Convolution integrated with SimVP), an efficient conditional spatiotemporal framework for global TEC forecasting built upon a purely convolutional SimVP backbone [8]. Instead of naïvely concatenating exogenous drivers, PMGC-SimVP encodes time-varying space weather conditions into a compact global context and injects it into multi-scale feature evolution via lightweight gated modulation, enabling region-sensitive local responses under strong non-stationarity. A residual prediction strategy over a simple baseline is further adopted to stabilize long-horizon forecasting while preserving computational efficiency.

Our contributions are summarized as follows:

- We formulate global TEC forecasting as *conditional spatiotemporal prediction*, where evolution depends on historical TEC states and time-varying space weather drivers, highlighting the need for *region-dependent* responses under identical forcing.
- We propose PMGC-SimVP, a purely convolutional predictor that integrates multi-scale gated feature evolution with parameter-conditioned modulation, tailored for global grids via Earth-aware padding and latitudinally structured modulation.

– Extensive experiments on global TEC forecasting and generic video benchmarks show that PMGC-SimVP improves accuracy and robustness, with clear gains under geomagnetic disturbances and long-horizon prediction.

2 Related Work

2.1 Empirical Ionospheric Models

TEC has long been modeled with empirical and semi-empirical approaches, notably the Klobuchar model [12], IRI [1], and the NeQuick family [20]. These models estimate TEC or ionospheric delay via low-dimensional parameterizations of external drivers (e.g., solar indices, geomagnetic location, local time, seasonality). However, simplified assumptions and compact parameterizations limit their ability to produce accurate global TEC grid forecasts, especially under disturbed geomagnetic conditions and across latitude-dependent regimes with strong regional variability [4].

2.2 Spatiotemporal Prediction Models

Spatiotemporal prediction is a core topic in computer vision. Early models are dominated by recurrent designs such as ConvLSTM variants [10, 17] and memory-augmented predictors (e.g., PredRNN family) for higher-order nonstationarity [34, 36–38], with related units improving temporal modeling via structural or motion-aware modules [3, 35]. To improve efficiency, convolutional predictors such as SimVP/SimVPv2 decouple spatial and temporal modeling [8, 29], while other lines explore lightweight temporal units or frequency-domain modeling (e.g., TAU, wavelets) [22, 30]. Transformers have also been used for Earth system forecasting but often scale poorly to long sequences [9]. Benchmarks such as OpenSTL enable systematic evaluation across models and protocols [31]. In this work, we focus on efficient convolutional predictors compatible with operational grids, and enhance them with structured conditional modulation.

2.3 Physics-aware and Conditional Modeling

Physics-aware learning is often pursued via physics-informed neural networks (PINNs), which impose governing equations as soft constraints but can be hard to optimize in high-dimensional, strongly non-stationary settings [11, 14, 24]. Instead of enforcing explicit physical laws, we incorporate physical drivers through lightweight conditional modulation that remains efficient for long-horizon spatiotemporal prediction.

Another line is conditional modeling, which adapts representations using exogenous variables. Representative techniques include FiLM, CondConv, and dynamic/conditional convolutions [5, 23, 33, 39]. However, most are developed for locally homogeneous planar data and do not explicitly handle strong spatial heterogeneity or global-scale structure, limiting direct use in global TEC

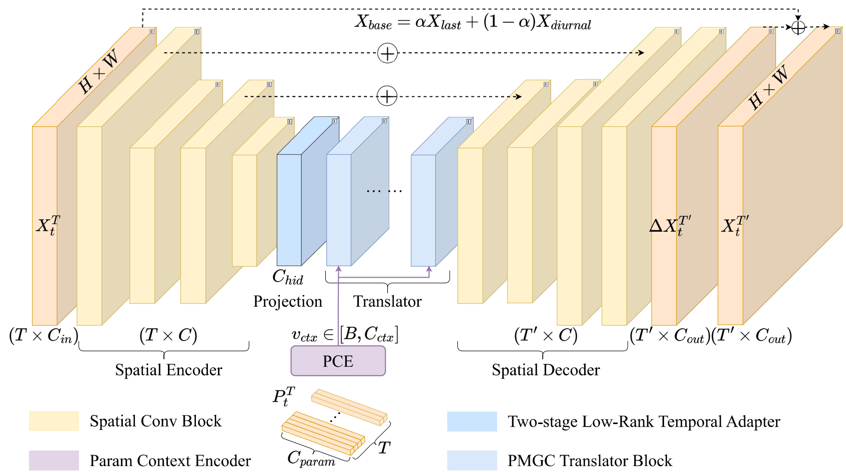

Fig. 1: Overview of the proposed PMGC-SimVP framework.

forecasting. In contrast, TEC exhibits pronounced latitudinal organization and region-dependent responses under identical forcing, motivating spatially structured conditional mechanisms.

3 Method

3.1 Problem Definition and Overall Framework

We study global ionospheric Total Electron Content (TEC) forecasting as a conditional spatiotemporal prediction problem. The TEC field is represented on a regular latitude–longitude grid, whose evolution is driven by external space weather processes and exhibits strong non-stationarity and pronounced latitudinal heterogeneity.

Formally, given a historical sequence $\mathcal{X}_{\text{in}} = \{X_t\}_{t=1}^{T_{\text{in}}}$, where $X_t \in \mathbb{R}^{C_{\text{in}} \times H \times W}$, the objective is to predict future states $\mathcal{X}_{\text{out}} = \{X_{T_{\text{in}}+t}\}_{t=1}^{T_{\text{out}}}$. System dynamics are further conditioned on an external parameter sequence $\mathcal{P}_{\text{in}} = \{p_t\}_{t=1}^{T_{\text{in}}}$, $p_t \in \mathbb{R}^{C_{\text{param}}}$, which provides global modulation signals for spatiotemporal evolution and corresponds to space weather drivers in TEC forecasting.

Based on this formulation, we propose PMGC-SimVP, a parameter modulated forecasting framework that separates globally stable structure from latitudinally structured, parameter-driven local dynamics via Earth-aware padding and band-wise modulation, as illustrated in Fig. 1.

In this work, we focus on an hourly-resolution forecasting setting, where the model takes the past $T_{\text{in}} = 48$ TEC maps as input and predicts the subsequent $T_{\text{out}} = 24$ maps. Given $\mathcal{X}_{\text{in}}$, the framework follows a SimVP-style encoder–projection–translator–decoder architecture. A spatial encoder extracts per-frame features, which are compressed by a two-stage low-rank temporal adapter into a compact latent representation $\mathbf{Z}$. The features are then processed by the PMGC translator, which incorporates parameter-modulated spatiotemporal evolution, and finally reconstructed by a spatial decoder to predict future TEC maps.

Parametric Context Encoding (PCE). The parameter sequence $\mathcal{P}_{\text{in}}$ is processed by PCE to extract time-varying space weather information. Using multi-query attention pooling, PCE produces a global context vector $\mathbf{v}_{\text{ctx}} \in \mathbb{R}^{B \times C_{\text{ctx}}}$, which is broadcast to all PMGC blocks for region-sensitive modulation.

3.2 Two-Stage Low-Rank Temporal Adapter

Latent features extracted by the encoder have high temporal dimensionality $(T_{\text{in}} \times C_{\text{hid}})$, making dense temporal projections inefficient for long sequences. We therefore introduce a two-stage low-rank temporal adapter that separates *channel-wise temporal modeling* from *cross-channel mixing*.

Each channel is first independently compressed from T_{in} to a temporal rank r_t using grouped temporal convolution ($g = C_{\text{hid}}$), followed by a pointwise convolution that mixes channels and projects them to a channel rank r_c. This reduces the complexity from $\mathcal{O}(T_{\text{in}} \cdot C_{\text{hid}}^2)$ to $\mathcal{O}(C_{\text{hid}} \cdot T_{\text{in}} \cdot r_t + C_{\text{hid}} \cdot r_t \cdot r_c)$.

With $T_{\text{in}} = 48$, $r_t = 12$, and $r_c = 64$, this design reduces parameters and FLOPs of the temporal adapter module by over 90%, exceeding 95% in practice, substantially improving long-sequence modeling efficiency.

3.3 PMGC Block

The PMGC Block is a plug-and-play unit that integrates global control parameters with local pixel-level dynamics (Fig. 2). It adopts a parallel *Slow–Fast* architecture to decouple global spatial structure from parameter-sensitive local dynamics.

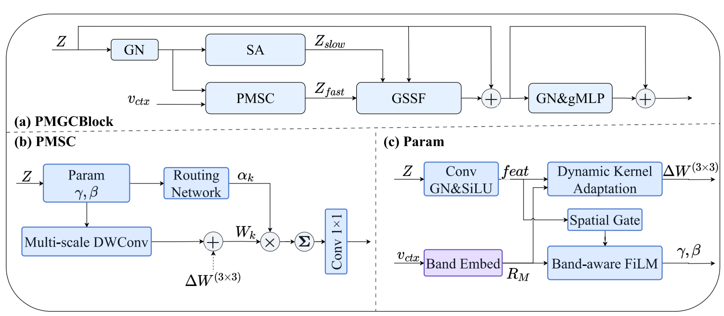

Fig. 2: Structure of the proposed PMGC Block. (a) Overall Slow–Fast PMGC block. (b) PMSC with adaptive routing and multi-scale depthwise convolution. (c) Param module for band-aware FiLM and dynamic kernel adaptation.

Slow Branch: Global Structure Modeling. The Slow branch captures global spatial structure using a large-kernel Spatial Aggregation (SA) module implemented via axial depthwise convolutions:

$$\mathbf{Z}_{\text{slow}} = \text{DWConv}_W\big(\text{DWConv}_H(\mathbf{Z})\big), \tag{1}$$

Fast Branch: Parametric Multi-Scale Local Modeling. The Fast branch models local dynamics using multi-scale depthwise convolutions with spatially adaptive routing. The input feature map to the Fast branch is defined as

$$\tilde{\mathbf{Z}} = \begin{cases} \mathbf{Z}, & \text{without parameter conditioning,} \\ \mathbf{Z}_{\text{mod}}, & \text{with parameter conditioning,} \end{cases} \tag{2}$$

$$\mathbf{Z}_{\text{fast}} = \sum_{k=1}^{K_S} \boldsymbol{\alpha}_k \odot \left(\mathbf{W}_k * \tilde{\mathbf{Z}}\right), \qquad \boldsymbol{\alpha} = \mathcal{R}(\tilde{\mathbf{Z}}), \tag{3}$$

where $\mathcal{R}$ denotes the routing network that predicts spatially adaptive weights over the multi-scale convolutional paths.

Band-wise Parametric Feature Modulation. To account for latitudinal heterogeneity, feature modulation is performed in a band-wise manner. Given the global context vector $\mathbf{v}_{\text{ctx}} \in \mathbb{R}^{B \times C_{\text{ctx}}}$ and band embeddings $\mathbf{e}_m$, we form band tokens

$$\mathbf{R}_M = \phi\Big(f\big(\text{Norm}(\mathbf{v}_{\text{ctx}} + \mathbf{e}_M)\big)\Big) \in \mathbb{R}^{B \times M \times r}, \tag{4}$$

which are broadcast to latitude rows to obtain $\mathbf{R} \in \mathbb{R}^{B \times H \times r}$. A content dependent spatial gate $\mathbf{A} \in \mathbb{R}^{B \times r \times H \times W}$ is computed from **feat**. Together with projection matrices $\mathbf{V}, \mathbf{B} \in \mathbb{R}^{r \times C}$, FiLM-style modulators are obtained as

$$\boldsymbol{\gamma}_{b,c,h,w} = \sum_{j=1}^{r} \mathbf{R}_{b,h,j}\, \mathbf{A}_{b,j,h,w}\, \mathbf{V}_{j,c}, \quad \boldsymbol{\beta}_{b,c,h,w} = \sum_{j=1}^{r} \mathbf{R}_{b,h,j}\, \mathbf{A}_{b,j,h,w}\, \mathbf{B}_{j,c}, \tag{5}$$

yielding the modulated feature map

$$\mathbf{Z}_{\text{mod}} = \mathbf{Z} \odot (1 + \boldsymbol{\gamma}) + \boldsymbol{\beta}. \tag{6}$$

Global Dynamic Kernel Adaptation. In addition to feature modulation, depthwise convolutional kernels are adapted using a global dynamic kernel mechanism. A prototype kernel bank $\mathcal{K} = \{\mathbf{K}_d\}_{d=1}^{D}$ ($\mathbf{K}_d \in \mathbb{R}^{3 \times 3}$) and a channel-wise projection $\mathbf{U} \in \mathbb{R}^{C \times D}$ are introduced. A unified global context is constructed as

$$\mathbf{u}_{\text{global}} = \text{Norm}\left(\frac{1}{M} \sum_{m=1}^{M} \mathbf{R}_{B,m,r} + \text{AttnPool}(\mathbf{feat})\right), \tag{7}$$

and used to predict path-specific prototype coefficients $\mathbf{a} \in \mathbb{R}^{B \times K_S \times D}$, where each convolutional path maintains an independent mixture over the shared prototype kernel bank. The effective kernel for the k-th scale is then

$$\mathbf{W}_{b,k,c} = \mathbf{W}_{k,c} + \sum_{d=1}^{D} a_{b,k,d}\, \mathbf{U}_{c,d}\, \mathbf{K}_d, \tag{8}$$

which is applied to $\mathbf{Z}_{\text{mod}}$.

Adaptive Fusion and Residual Update. The outputs of the two branches are fused with $\mathbf{Z}$ by a Grouped Spatial Softmax Fusion (GSSF) module, followed by a residual update, a lightweight gMLP, and a second residual refinement.

3.4 Dual-Baseline Residual Prediction

To mitigate error accumulation in long-horizon forecasting, we adopt a *Dual-Baseline Residual Learning* strategy. This decomposes the prediction into a deterministic baseline term and a network-predicted residual:

$$\hat{\mathcal{X}}^{(t)} = \mathcal{X}^{(t)}_{\text{base}} + \Delta\mathcal{X}^{(t)}, \tag{9}$$

where the residual $\Delta\mathcal{X}^{(t)}$ captures non-stationary dynamics. The baseline $\mathcal{X}^{(t)}_{\text{base}}$ combines short-term inertia and long-term periodicity:

$$\mathcal{X}^{(t)}_{\text{base}} = \alpha_t \cdot X_{\text{last}} + (1 - \alpha_t) \cdot X^{(t)}_{\text{diurnal}}. \tag{10}$$

Here, X_{last} is the most recent observation, and $X^{(t)}_{\text{diurnal}}$ denotes the diurnal reference. The weight α_t decreases linearly from 1 to 0 over the forecast horizon, smoothly transitioning the prior from short-term continuity to long-term periodic trends.

3.5 Objective Function

Masked Charbonnier Loss. To mitigate boundary discontinuities induced by longitude wrap-around, we employ Earth-aware padding (circular in longitude, replicate in latitude) in the encoder to maintain spatial continuity. Due to preprocessing and padding, a fixed validity mask is applied during training, and the loss is computed exclusively over physically valid grid locations:

$$\mathcal{L} = \frac{1}{T_{\text{out}}\, |\Omega_{\text{valid}}|} \sum_{t=1}^{T_{\text{out}}} \sum_{(h,w)\in\Omega_{\text{valid}}} \sqrt{\Big(\hat{\mathcal{X}}_t(h,w) - \mathcal{X}_t(h,w)\Big)^2 + \epsilon^2}. \tag{11}$$

Table 1: Task settings and dataset statistics.

Dataset	Train	Test	C	H	W	T_{in}	T_{out}
TEC	13130	727	1	72	72	48	24
WeatherBench	2167	706	1	32	64	12	12
TaxiBJ	20461	500	2	32	32	4	4

4 Experiments

4.1 Experimental Settings

Datasets and Protocols. We evaluate our method on two types of tasks, with dataset statistics and protocol settings summarized in Table 1.

(i) Global TEC forecasting. We use hourly global IGS TEC maps (71×73), which are converted to a unified 72×72 grid by removing the duplicated longitude and applying latitude-wise replicate padding. The dataset includes external space weather parameters, including F10.7, SSN, $\mathbf{W}_{\text{low}}$, $\mathbf{W}_{\text{high}}$, C, M, X, Dst, X-ray, and Ap, together with the corresponding timestamp (*year, month, day, hour*).

(ii) Generic benchmarks. WeatherBench [25] and TaxiBJ [40] are evaluated under the *w/o param* setting, following the OpenSTL [31] protocol.

Baselines and Metrics. During training, the loss is computed on a fixed 71×72 valid subregion by excluding the padded latitude row, avoiding boundary artifacts. For evaluation, predictions are mapped back to the restored 71×73 grid. We report MAE, RMSE, SSIM, Pearson correlation (Corr.), and gradient RMSE (Grad.).

Corr. is computed as the global Pearson correlation coefficient over all valid spatiotemporal elements. Grad. measures the isotropic RMSE of first-order spatial gradients, averaged over longitude and latitude directions, with masks applied where appropriate.

Parameter importance is quantified on the input parameter vector using gradient–input attribution:

$$\mathcal{I} = \mathbb{E}_{(x,p),t}\left[\left|\frac{\partial \mathcal{L}}{\partial p_t} \odot p_t\right|\right], \tag{12}$$

where $\mathcal{L}$ denotes the prediction loss and p_t the input parameter at time t.

Implementation Details. All models are trained using AdamW (lr$=5\times10^{-4}$, batch size 16) on a single NVIDIA RTX 4090D GPU. For WeatherBench and TaxiBJ, we adopt the *w/o param* setting and follow OpenSTL [31] defaults. All baselines are trained and evaluated under the OpenSTL protocol with the same TEC inputs/targets, preprocessing, and valid-region mask.

Table 2: Comparison of TEC prediction performance under stormy conditions in 2023 and the Overall Average across test periods. See supplementary for details.

Method	Year 2023 (Stormy)			Overall Average		
	MAE ↓	RMSE ↓	SSIM ↑	MAE ↓	RMSE ↓	SSIM ↑
SimVP [8]	3.257	4.815	0.933	2.043	3.515	0.959
PredRNN-v2 [37]	3.568	5.196	0.922	2.256	3.809	0.952
TAU [30]	3.359	4.826	0.930	2.119	3.532	0.957
WaST [22]	2.986	4.439	0.940	1.882	3.248	0.964
Stormer [21]	3.133	4.581	0.933	1.978	3.357	0.959
Met2Net [16]	3.028	4.500	0.941	1.896	3.288	0.964
Ours	**2.801**	**4.159**	**0.946**	**1.774**	**3.051**	**0.967**

Table 3: Corr. and Grad. for assessing temporal consistency and structural preservation across latitude bands under stormy conditions in 2023.

Method	Low Lat.		Mid Lat.		High Lat.	
	Corr. ↑	Grad. ↓	Corr. ↑	Grad. ↓	Corr. ↑	Grad. ↓
SimVP [8]	0.971	1.037	0.963	0.622	0.909	0.468
PredRNN-v2 [37]	0.965	1.224	0.955	0.729	0.896	0.523
TAU [30]	0.970	1.025	0.959	0.642	0.908	0.486
WaST [22]	0.975	0.969	0.968	0.577	0.921	0.443
Stormer [21]	0.973	1.206	0.965	0.723	0.919	0.553
Met2Net [16]	0.975	0.961	0.967	0.582	0.917	0.451
Ours	**0.978**	**0.914**	**0.971**	**0.553**	**0.930**	**0.428**

4.2 Main Results and Robustness Analysis

Table 2 reports quantitative comparisons on global TEC forecasting. On average across all test periods, PMGC-SimVP achieves the best performance across MAE, RMSE, and SSIM. Compared with the strongest baseline WaST, our method reduces RMSE from 3.248 to 3.051 and improves SSIM to 0.967, indicating consistent advantages in both numerical accuracy and spatial structure preservation.

Performance under Stormy Conditions (2023). The year 2023 lies in the ascending phase of Solar Cycle 25, where frequent geomagnetic storms intensify non-stationary TEC dynamics. As shown in Table 2, all baselines suffer notable degradation under stormy conditions, whereas PMGC-SimVP maintains a lower RMSE of 4.159. Notably, the relative performance gap between PMGC-SimVP and the baselines widens during storm periods, suggesting improved robustness under strong space-weather forcing.

Structural and Temporal Preservation. To assess spatiotemporal consistency during high-activity periods, we report Corr. and Grad. for 2023 in Table 3. PMGC-SimVP consistently achieves the best Corr. and Grad. across low-, mid-, and high-latitude regions, e.g., Corr. = 0.978 and Grad. = 0.914 in low latitudes.

Low latitudes contain the Equatorial Ionization Anomaly (EIA), which has a sharp gradient, making it particularly challenging to predict during intense space weather events. The consistent improvements in Corr. and Grad. indicate that the proposed latitudinal band-wise modulation effectively captures region-dependent dynamics and regularizes high-gradient EIA structures during storms.

4.3 Long-Horizon Stability and Regime Adaptation

Fig. 3 reports frame-wise MAE over a 24-hour forecast horizon under different geomagnetic activity regimes. While all methods exhibit diurnal error patterns, baselines show persistently higher errors and larger temporal fluctuations

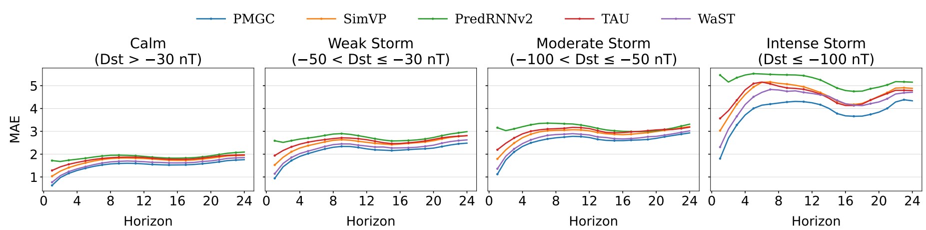

Fig. 3: Frame-wise MAE over a 24-hour forecast horizon (Overall Average).

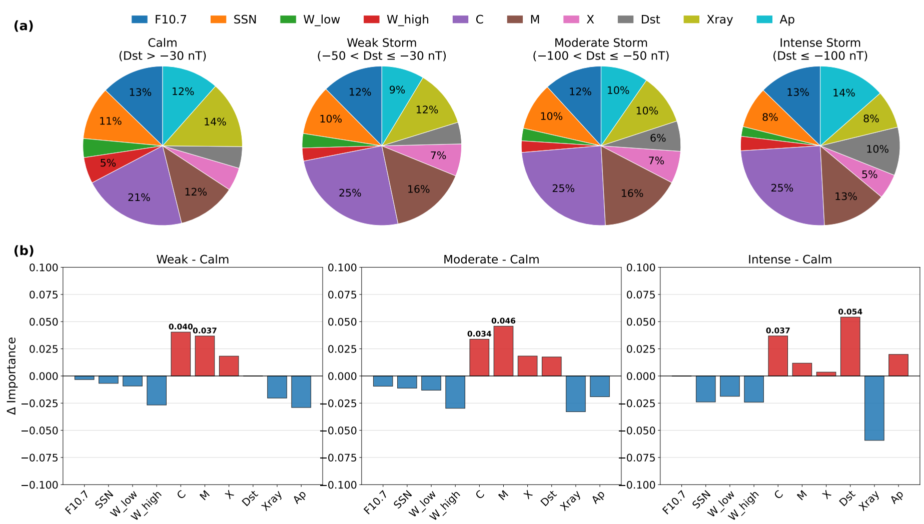

Fig. 4: Physical parameter importance (stratified by geomagnetic conditions). (a) Relative composition. (b) Changes with respect to Calm conditions.

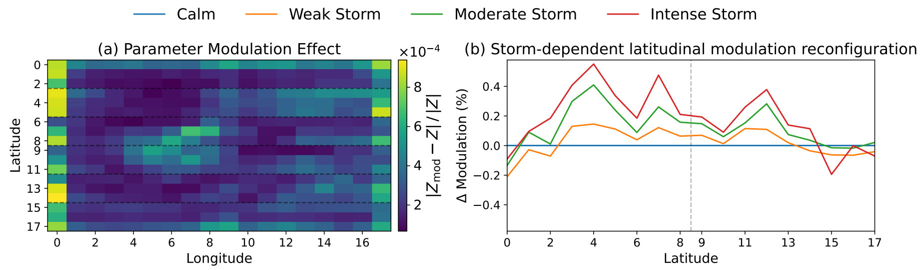

Fig. 5: Latitudinal patterns of parameter modulation effect. (a) Spatial pattern of modulation effect. (b) Storm-dependent latitudinal modulation reconfiguration.

as storm intensity increases. In contrast, PMGC-SimVP generally maintains a lower and flatter error profile over most horizons, with a widening performance gap in moderate and intense storm conditions, indicating improved long-horizon stability.

To interpret the source of long-horizon stability, Fig. 4 presents gradient–input attribution of space weather parameters across geomagnetic regimes. Under *Calm* conditions, attribution is dominated by radiative and background descriptors, notably the Number of C flares, along with X-ray, F10.7, Ap, and M. As geomagnetic activity increases from Weak to Moderate, importance is reallocated rather than uniformly amplified: C increases monotonically, while background indices such as SSN, F10.7, and Ap exhibit a net reduction relative to Calm. Under *Intense* conditions, a clear transition emerges. Although C remains dominant, Dst exhibits the largest positive shift and a substantially increased share, indicating stronger reliance on the accumulated magnetospheric state. Meanwhile, Ap regains importance compared to weaker storms, suggesting renewed discriminative power under extreme disturbances.

To further examine how parameter conditioning manifests spatially, Fig. 5 visualizes the relative modulation effect and its storm-dependent latitudinal variation. As shown in Fig. 5, parameter-induced modulation exhibits a clear latitudinal structure, with stronger responses concentrated in low-latitude regions associated with the EIA. Relative to Calm conditions, modulation strength increases non-uniformly as geomagnetic activity intensifies, reaching the largest amplification under Intense storms. Importantly, this enhancement is not a uniform global scaling, but a regime-dependent reconfiguration of latitudinal modulation. These results suggest that PMGC adapts parameter influence in a spatially structured manner, enabling external drivers to selectively modulate storm-sensitive regions.

4.4 Efficiency Analysis

Table 4 compares model complexity and inference efficiency. SimVP, TAU, and ours use the same configuration; others follow OpenSTL or original reports.

Table 4: Model complexity and efficiency comparison (measured on RTX 4090D).

Method	Params (M) ↓	FLOPs (G) ↓	FPS ↑	Mem (GB)
SimVP [8]	195	91.71	989.8	10.72
PredRNN-v2 [37]	23.86	551	261.5	5.17
TAU [30]	186	88.59	473.1	10.60
WaST [22]	142	175	344.0	18.15
Ours	**9.89**	**19.905**	**1745.2**	5.97

Benefiting from the two-stage low-rank temporal adapter, PMGC-SimVP contains only 9.89M parameters, corresponding to 5.07% of SimVP (195M). By avoiding RNN-style sequential computation, PMGC-SimVP achieves an inference throughput of 1745.2 fps, outperforming SimVP (989.8 fps) and yielding a 5.07× speedup over WaST (344.0 fps). These results demonstrate that PMGC-SimVP attains substantially higher inference efficiency while maintaining low computational and memory overhead under identical hardware and implementation settings.

4.5 Ablation Studies

We conduct systematic ablation studies to evaluate the contribution of individual components, with results summarized in Table 5. Removing any module consistently degrades performance, with more pronounced effects in low-latitude regions.

Among architectural ablations, removing parameter condition (w/o Param Condition) leads to consistent degradation across all latitude bands, particularly in low latitudes. Eliminating the PMSC module (w/o PMSC), the SA module

Table 5: Ablation study of component effectiveness. We report MAE by latitude region and overall RMSE. "w/o" removes a module. Param=0 zeros input parameters, and Param-Shuffled permutes them along the batch dimension, with the network fixed.

Variants	Low Lat.	Mid Lat.	High Lat.	Overall Average	
	MAE ↓	MAE ↓	MAE ↓	MAE ↓	RMSE ↓
w/o Charbonnier	2.497	1.530	1.536	1.845	3.164
w/o Dual-Baseline Residual	2.530	1.503	1.528	1.844	3.161
w/o Param Condition	2.462	1.480	1.500	1.804	3.118
w/o PMSC	2.455	1.481	1.516	1.808	3.132
w/o SA	2.455	1.475	1.505	1.802	3.092
Ours (Param=0)	–	–	–	1.823	3.074
Ours (Param-Shuffled)	–	–	–	1.784	3.066
Ours (Full)	**2.426**	**1.453**	**1.471**	**1.774**	**3.051**

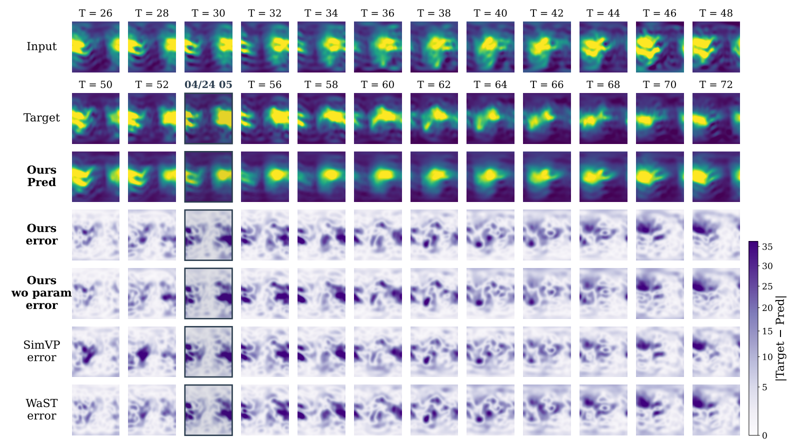

Fig. 6: Qualitative comparison during the April 23–24, 2023 severe geomagnetic superstorm. Top to bottom: Input, GT, Ours, and error maps.

(w/o SA), or the dual-baseline residual (w/o Dual-Baseline Residual) also results in noticeable performance drops, indicating their complementary roles in spatial modeling and long-horizon stabilization.

Beyond module removal, we include two controlled variants that perturb only the parameter inputs while keeping the model architecture and weights unchanged. Both Param=0 and Param-Shuffled exhibit clear performance degradation compared with the full model, with setting parameters to zero causing the largest drop.

Overall, the ablation results demonstrate that parameter-conditioned routing, multi-scale local modeling, and dual-baseline residual prediction address complementary aspects of TEC dynamics. Their synergy, rather than any single component, is crucial for the final performance.

4.6 Visualization of Severe Geomagnetic Superstorm

To evaluate model behavior under extreme space weather conditions, Fig. 6 presents a qualitative comparison during the April 23–24, 2023 severe geomagnetic superstorm, which occurred in the ascending phase of Solar Cycle 25 with a peak Dst of approximately -213 nT.

This event exhibits a pronounced regime shift in the ground-truth TEC field: EIA crests rapidly weaken and reorganize, consistent with a storm-time *negative phase* characterized by large-scale TEC depletion. Such dynamics deviate substantially from historical and diurnal patterns, posing a severe challenge for long-horizon forecasting.

As shown in Fig. 6, PMGC-SimVP outperforms SimVP and WaST during the early and middle stages of the event, capturing the initial magnitude suppression and structural reorganization. Beyond approximately $T = 67$ h, this advantage

Table 6: Generalization performance of the PMGC backbone on non-TEC benchmarks (w/o param). Baseline results are taken from the corresponding original papers or OpenSTL benchmarks. Metrics not applicable to a dataset are marked as "–".

Method	WeatherBench			TaxiBJ			
	MSE ↓	**MAE ↓**	**RMSE ↓**	**MSE ↓**	**MAE ↓**	**SSIM**	**PSNR ↑**
ConvLSTM [17]	1.521	0.795	1.233	48.5	17.7	0.978	37.38
PredRNN [36]	1.331	0.725	1.154	46.4	17.1	0.971	38.52
PredRNN++ [34]	1.634	0.788	1.278	44.8	16.9	0.977	38.71
MIM [38]	1.784	0.872	1.336	42.9	16.6	0.971	38.71
E3D-LSTM [35]	1.592	0.806	1.233	43.2	16.9	0.979	38.75
MAU [3]	1.349	0.739	1.162	–	–	–	–
SimVP [8]	1.238	0.704	1.113	41.4	16.2	0.982	39.17
PredRNN-v2 [37]	1.545	0.799	1.243	–	–	–	–
TAU [30]	1.224	0.681	1.106	34.4	15.6	0.983	39.50
WaST [22]	1.098	0.634	1.044	30.8	14.9	0.984	39.73
Ours	**1.026**	**0.612**	**1.013**	**28.3**	**14.5**	0.980	**39.88**

diminishes, and the prediction error becomes comparable to or slightly larger than that of the baselines. This behavior arises because the model increasingly relies on long-term periodic trends at later prediction stages, while the ionosphere remains in an exceptionally strong and prolonged negative storm phase, causing the deviation from historical references to further increase.

Importantly, this degradation is not systematic. As shown in the supplementary analysis of another intense storm (Dst $= -163$ nT), PMGC-SimVP retains lower errors over most forecast horizons, indicating improved robustness under strong geomagnetic disturbances in general.

4.7 Generalization

We evaluate the proposed backbone on WeatherBench and TaxiBJ to assess cross-task generalization *without explicit external control parameters*. Since neither dataset provides exogenous variables in OpenSTL, we adopt the w/o param setting, where predictions are formulated as residuals over the last observed frame. Accordingly, PMGC-SimVP reduces to a multi-scale gated convolutional backbone with residual modeling, and the results primarily reflect its structured spatiotemporal modeling capability rather than parameter modulation.

As shown in Table 6, the proposed method achieves the lowest MSE on WeatherBench and the best MSE, MAE, and PSNR on TaxiBJ, while maintaining comparable SSIM. Despite substantial differences in dynamics, grid topology, and data semantics, these results demonstrate strong cross-task generalization of the proposed backbone. For systems with explicit external drivers (e.g., TEC forecasting), the parametric modulation mechanism can be naturally integrated on top of this robust backbone to further improve performance.

5 Conclusion

This paper proposes PMGC-SimVP, a parametric multi-scale gated framework for global ionospheric TEC forecasting under strong non-stationarity. PMGC-SimVP integrates space weather drivers into a convolutional predictor by decoupling global structure from parameter-sensitive local dynamics, together with Earth-aware padding and latitudinally structured modulation.

Experiments demonstrate that PMGC-SimVP achieves improved overall performance compared with recent strong baselines, with clear gains during geomagnetic disturbances and in complex low-latitude regions. Ablation studies confirm the complementary effects of parametric modulation, multi-scale local modeling, and dual-baseline residual prediction in stabilizing long-horizon forecasting. Notably, the proposed backbone maintains competitive performance even without parameter injection, indicating the intrinsic effectiveness of the structural designs.

Beyond TEC, the proposed backbone suggests a practical framework for conditional spatiotemporal prediction of global geophysical fields: control signals can be integrated when available, while the backbone remains effective without them. Future work will explore finer-grained driver representations and extend the framework to higher-resolution fields and multi-source forecasting.

References

1. Bilitza, D., Pezzopane, M., Truhlik, V., Altadill, D., Reinisch, B.W., Pignalberi, A.: The international reference ionosphere model: A review and description of an ionospheric benchmark. Reviews of geophysics **60**(4), e2022RG000792 (2022)
2. Camporeale, E.: The challenge of machine learning in space weather: Nowcasting and forecasting. Space weather **17**(8), 1166–1207 (2019)
3. Chang, Z., Zhang, X., Wang, S., Ma, S., Ye, Y., Xinguang, X., Gao, W.: Mau: A motion-aware unit for video prediction and beyond. Advances in Neural Information Processing Systems **34**, 26950–26962 (2021)
4. Chen, J., Ren, X., Zhang, X., Zhang, J., Huang, L.: Assessment and validation of three ionospheric models (iri-2016, nequick2, and igs-gim) from 2002 to 2018. Space Weather **18**(6), e2019SW002422 (2020)
5. Chen, Y., Dai, X., Liu, M., Chen, D., Yuan, L., Liu, Z.: Dynamic convolution: Attention over convolution kernels. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11030–11039 (2020)
6. Cohen, T.S., Geiger, M., Koehler, J., Welling, M.: Spherical cnns (2018), `https://arxiv.org/abs/1801.10130`
7. Defferrard, M., Milani, M., Gusset, F., Perraudin, N.: Deepsphere: a graph-based spherical cnn. In: International Conference on Learning Representations (2020), `https://openreview.net/forum?id=B1e3OlStPB`
8. Gao, Z., Tan, C., Wu, L., Li, S.Z.: Simvp: Simpler yet better video prediction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3170–3180 (2022)
9. Gao, Z., Shi, X., Wang, H., Zhu, Y., Wang, Y.B., Li, M., Yeung, D.Y.: Earthformer: Exploring space-time transformers for earth system forecasting. Advances in Neural Information Processing Systems **35**, 25390–25403 (2022)

10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
11. Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L.: Physics-informed machine learning. Nature Reviews Physics **3**(6), 422–440 (2021)
12. Klobuchar, J.A.: Ionospheric time-delay algorithm for single-frequency gps users. IEEE Transactions on Aerospace and Electronic Systems **AES-23**(3), 325–331 (May 1987). https://doi.org/10.1109/TAES.1987.310829
13. Krachmalnicoff, N., Tomasi, M.: Convolutional neural networks on the healpix sphere: a pixel-based algorithm and its application to cmb data analysis. Astronomy & Astrophysics **628**, A129 (2019)
14. Krishnapriyan, A., Gholami, A., Zhe, S., Kirby, R., Mahoney, M.W.: Characterizing possible failure modes in physics-informed neural networks. Advances in neural information processing systems **34**, 26548–26560 (2021)
15. Lean, J., Meier, R., Picone, J., Sassi, F., Emmert, J., Richards, P.: Ionospheric total electron content: Spatial patterns of variability. Journal of Geophysical Research: Space Physics **121**(10), 10–367 (2016)
16. Li, S., et al.: Met2Net: A decoupled two-stage spatio-temporal forecasting model for complex meteorological systems. In: ICCV. pp. 5458–5468 (2025)
17. Lin, Z., Li, M., Zheng, Z., Cheng, Y., Yuan, C.: Self-attention convlstm for spatiotemporal prediction. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 11531–11538 (2020)
18. Liu, L., Zou, S., Yao, Y., Wang, Z.: Forecasting global ionospheric tec using deep learning approach. Space Weather **18**(11), e2020SW002501 (2020)
19. Luo, H., Gong, Y., Chen, S., Yu, C., Yang, G., Yu, F., Hu, Z., Tian, X.: Prediction of global ionospheric total electron content (TEC) based on SAM-ConvLSTM model. Space Weather **21**(12) (2023). https://doi.org/https://doi.org/10.1029/2023SW003707
20. Nava, B., Coisson, P., Radicella, S.: A new version of the nequick ionosphere electron density model. Journal of atmospheric and solar-terrestrial physics **70**(15), 1856–1862 (2008)
21. Nguyen, T., et al.: Scaling transformer neural networks for skillful and reliable medium-range weather forecasting. arXiv:2312.03876 (2023)
22. Nie, X., Yan, Y., Li, S., Tan, C., Chen, X., Jin, H., Zhu, Z., Li, S.Z., Qi, D.: Wavelet-driven spatiotemporal predictive learning: Bridging frequency and time variations. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 4334–4342 (2024)
23. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
24. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational physics **378**, 686–707 (2019)
25. 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) (Nov 2020). https://doi.org/10.1029/2020ms002203, http://dx.doi.org/10.1029/2020MS002203
26. Ren, X., Yang, P., Liu, H., Chen, J., Liu, W.: Deep learning for global ionospheric tec forecasting: Different approaches and validation. Space Weather **20**(5), e2021SW003011 (2022)

27. Ruwali, A., Kumar, A.S., Prakash, K.B., Sivavaraprasad, G., Ratnam, D.V.: Implementation of hybrid deep learning model (lstm-cnn) for ionospheric tec forecasting using gps data. IEEE Geoscience and Remote Sensing Letters **18**(6), 1004–1008 (2020)
28. Schaer, S., helvétique des sciences naturelles. Commission géodésique, S.: Mapping and predicting the Earth's ionosphere using the Global Positioning System, vol. 59. Institut für Geodäsie und Photogrammetrie, Eidg. Technische Hochschule ... (1999)
29. Tan, C., Gao, Z., Li, S., Li, S.Z.: Simvpv2: Towards simple yet powerful spatiotemporal predictive learning. IEEE Transactions on Multimedia (2025)
30. Tan, C., Gao, Z., Wu, L., Xu, Y., Xia, J., Li, S., Li, S.Z.: Temporal attention unit: Towards efficient spatiotemporal predictive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 18770–18782 (2023)
31. Tan, C., Li, S., Gao, Z., Guan, W., Wang, Z., Liu, Z., Wu, L., Li, S.Z.: Openstl: A comprehensive benchmark of spatio-temporal predictive learning. In: Conference on Neural Information Processing Systems Datasets and Benchmarks Track (2023)
32. Tapping, K.: The 10.7 cm solar radio flux (f10. 7). Space weather **11**(7), 394–406 (2013)
33. Tian, Z., Shen, C., Chen, H.: Conditional convolutions for instance segmentation. In: European conference on computer vision. pp. 282–298. Springer (2020)
34. Wang, Y., Gao, Z., Long, M., Wang, J., Yu, P.S.: Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In: International conference on machine learning. pp. 5123–5132. PMLR (2018)
35. Wang, Y., Jiang, L., Yang, M.H., Li, L.J., Long, M., Fei-Fei, L.: Eidetic 3d lstm: A model for video prediction and beyond. In: International Conference on Learning Representations (2018)
36. Wang, Y., Long, M., Wang, J., Gao, Z., Yu, P.S.: Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. Advances in neural information processing systems **30** (2017)
37. Wang, Y., Wu, H., Zhang, J., Gao, Z., Wang, J., Yu, P.S., Long, M.: Predrnn: A recurrent neural network for spatiotemporal predictive learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(2), 2208–2225 (2022)
38. Wang, Y., Zhang, J., Zhu, H., Long, M., Wang, J., Yu, P.S.: Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9154–9162 (2019)
39. Yang, B., Bender, G., Le, Q.V., Ngiam, J.: Condconv: Conditionally parameterized convolutions for efficient inference. Advances in neural information processing systems **32** (2019)
40. Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proceedings of the AAAI conference on artificial intelligence. vol. 31 (2017)