

VeinCast: Physics-Guided Dynamic Field Graphs with Graph-Conditioned Fusion for Global Medium-Range Weather Forecasting

Zhisheng Chen^{1,*}, Jinhan Li^{1,*}, Yuxuan Li¹, Yuan Gao², Hao Wu², Zheng Lu³, Jinlong Du^{4,†}, Kun Wang^{1,†}, Bo An¹

¹Nanyang Technological University ²Tsinghua University ³Peking University ⁴Tongji University

* Equal contribution † Corresponding authors ✉ zhisheng.researcher@gmail.com

🔗 <https://github.com/Zhisheng-researcher/VeinCast>

Abstract Global medium-range weather forecasting requires modeling structured yet state-dependent interactions among heterogeneous atmospheric fields. Existing data-driven models largely learn these interactions implicitly, whereas equation-level physical constraints may inherit approximation and model-form biases. We present VeinCast, a physics-guided dynamic field graph and graph-conditioned fusion framework that jointly forecasts 69 surface and upper-air fields. Within each local window, its Physics-Guided Dynamic Field Graph combines predefined atmospheric relations with state-dependent Top-K residual edges and adapts Earth-window attention using the resulting graph context. Graph-Conditioned Latent Fusion further employs graph context and source-node centrality to guide field-to-latent aggregation, while bounded feedback preserves field-specific information. On the 1.5° ERA5 benchmark, VeinCast demonstrates competitive forecasting performance across all 69 meteorological fields at lead times of up to 14 days, compared with representative global weather forecasting models including FuXi, Pangu-Weather, GraphCast, FengWu, and ARROW. Ablations confirm that the two modules provide complementary gains, demonstrating the effectiveness of relational-level physical guidance for data-driven weather forecasting.

Introduction

Medium-range weather forecasting predicts atmospheric evolution over lead times from several days to approximately two weeks and supports weather-sensitive decisions in disaster preparedness, agriculture, energy, transportation, and water-resource management [Price et al. \(2025\)](#); [Zhao et al. \(2024\)](#). Reliable global forecasting requires jointly modeling heterogeneous surface and upper-air fields across multiple vertical levels [Chen et al. \(2023b\)](#). Their interactions exhibit persistent physical structure, but their relevance and strength vary across locations, pressure levels, and atmospheric states [Al-Yahyai, Charabi, and Gastli \(2010\)](#). Capturing these structured yet state-dependent cross-field couplings is therefore essential for skillful and computationally efficient medium-range forecasting.

Operational forecasting traditionally relies on Numerical Weather Prediction (NWP), which uses data assimilation to initialize the atmosphere and numerically integrates discretized governing equations [Bauer, Thorpe, and Brunet \(2015\)](#). Although physically grounded, high-resolution integration and ensemble forecasting are computationally demanding, while errors from finite resolution, unresolved-process parameterizations, numerical discretization, and imperfect initial conditions accumulate with lead time [Allen et al. \(2025\)](#). These trade-offs have motivated data-driven approaches that learn atmospheric state transitions from historical data [Maraun et al. \(2010\)](#).

Recent AI forecasters have established a competitive data-driven paradigm for global weather forecasting while substantially reducing forecast-generation costs [Bi et al. \(2023\)](#); [Chen et al. \(2023a\)](#); [Chen, Zhong, and Zhang \(2023\)](#); [Lam et al. \(2023\)](#); [Pathak et al. \(2022\)](#); [Ren et al. \(2021\)](#). Nevertheless, many architectures represent meteorological fields as channels or tokens and infer their interactions

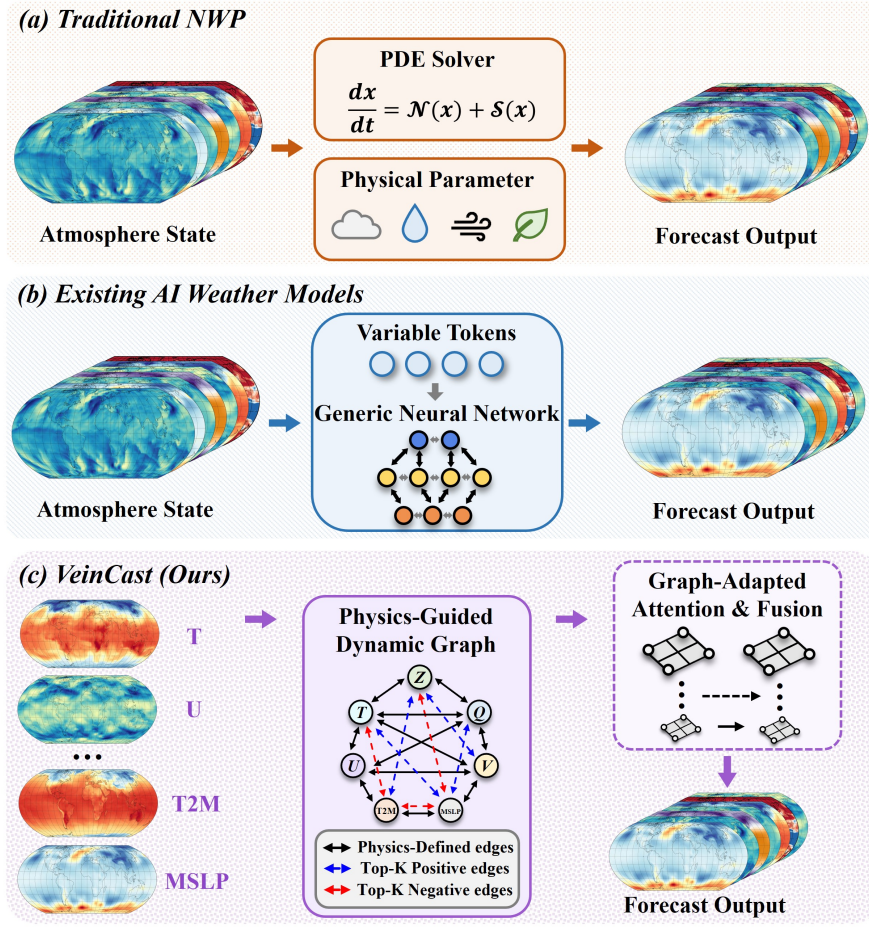


Figure 1 | Overview of weather forecasting paradigms and VeinCast. **(a)** Traditional NWP relies on numerical solvers and physical parameterizations. **(b)** Existing AI models process atmospheric variables using generic neural networks. **(c)** VeinCast models structured variable interactions through a physics-guided dynamic field graph and graph-conditioned latent fusion.

through generic convolution, attention, or message-passing operators [Gan et al. \(2025\)](#). Graph-based forecasters likewise construct graphs primarily over spatial grid or mesh nodes, leaving cross-field dependencies implicit in feature representations [Gao et al. \(2025\)](#). Such designs can learn statistical couplings but do not explicitly distinguish persistent atmospheric relations from interactions whose relevance changes with the local state. Vertical dependencies and couplings among wind, geopotential, temperature, and moisture have recognizable physical structure, yet their effective strengths vary across atmospheric conditions.

Physics–AI hybrid models offer another route by embedding governing equations, differentiable solvers, or equation-based constraints into neural forecasting [Kochkov et al. \(2024\)](#); [Xu et al. \(2025\)](#). Although these approaches provide strong physical guidance, the incorporated formulations may still reflect approximations, closure assumptions, parameterizations, and numerical discretization choices. Imposing such formulations too rigidly may transfer model-form biases to the learning process and restrict data-driven correction, while remaining errors can be amplified during chaotic forecast rollout. This motivates incorporating physical knowledge at the relational rather than equation level: selected relationships provide a stable inductive bias without prescribing exact field evolution. However, a fixed relation graph cannot describe all regime-dependent interactions, whereas learning all relations without guidance may overlook robust atmospheric knowledge. We therefore combine a

physics-defined backbone with data-dependent connections that capture couplings emerging under specific local states. The resulting graph should also guide multiscale spatial modeling and cross-field aggregation rather than remain an isolated auxiliary representation, while avoiding excessive fusion that obscures field-specific characteristics.

To address these challenges, we propose VeinCast, a physics-guided dynamic field graph and graph-conditioned fusion framework for global medium-range weather forecasting. VeinCast learns a 6-hour state transition and generates medium-range forecasts through autoregressive rollout. We consider a closed-set setting with 69 input and output fields, including 65 upper-air fields from five variables at 13 predefined pressure levels and four surface fields. Within each local spatial window, these fields form the nodes of a directed graph. Physics-defined edges provide a stable relational scaffold, while state-dependent Top- K residual edges capture additional couplings absent from the predefined relations. The resulting graph context adapts Earth-window spatial attention and conditions field-to-latent aggregation through source-node centrality. A Latent U-Backbone models multiscale spatial dependencies and interactions among fusion latents, after which bounded feedback returns the fused context without overwhelming field-specific features. Finally, the Forecast Decoder reconstructs the fixed forecast fields as residual atmospheric updates.

In summary, our main contributions are as follows:

- We introduce a Physics-Guided Dynamic Field Graph that incorporates physical knowledge at the relational rather than equation level. It combines a stable physics-defined scaffold to distinguish persistent atmospheric relations from locally varying couplings, while using the graph context to adapt Earth-window spatial attention.
- We develop Graph-Conditioned Latent Fusion, which uses dynamic graph context and source-node centrality to guide field-to-latent aggregation. A multiscale latent backbone captures shared interactions, while bounded feedback returns the fused context without obscuring field-specific information.
- Systematic experiments on global medium-range forecasting demonstrate competitive accuracy across representative surface and upper-air fields. Ablation studies further verify the complementary contributions of the dynamic field graph and graph-conditioned fusion.

Related Work

Numerical Weather Prediction.

NWP initializes the atmosphere through data assimilation and advances it by numerically solving discretized governing equations [Molteni et al. \(1996\)](#). Although physically grounded, high-resolution and ensemble forecasting remain computationally expensive and are affected by initial-condition, resolution, parameterization, and discretization errors [Nguyen et al. \(2023\)](#); [Wedi et al. \(2015\)](#).

Data-Driven Weather Forecasting.

Recent advances in deep learning and large-scale reanalysis datasets have accelerated the development of data-driven weather forecasting. Representative Transformer-based models include Pangu-Weather, FengWu, FuXi, Stormer [Nguyen et al. \(2024\)](#), STCast [Chen et al. \(2026a\)](#), EMFormer [Chen et al. \(2026b\)](#), and Aurora [Bodnar et al. \(2025\)](#). Graph-based approaches include Keisler’s model [Keisler \(2022\)](#), GraphCast, OneForecast, and Graph-EFM [Oskarsson et al. \(2024\)](#), while diffusion-based methods include GenCast and FengWu-ENS [Chen et al. \(2025\)](#).

Beyond purely data-driven approaches, physics–AI hybrid methods incorporate physical mechanisms

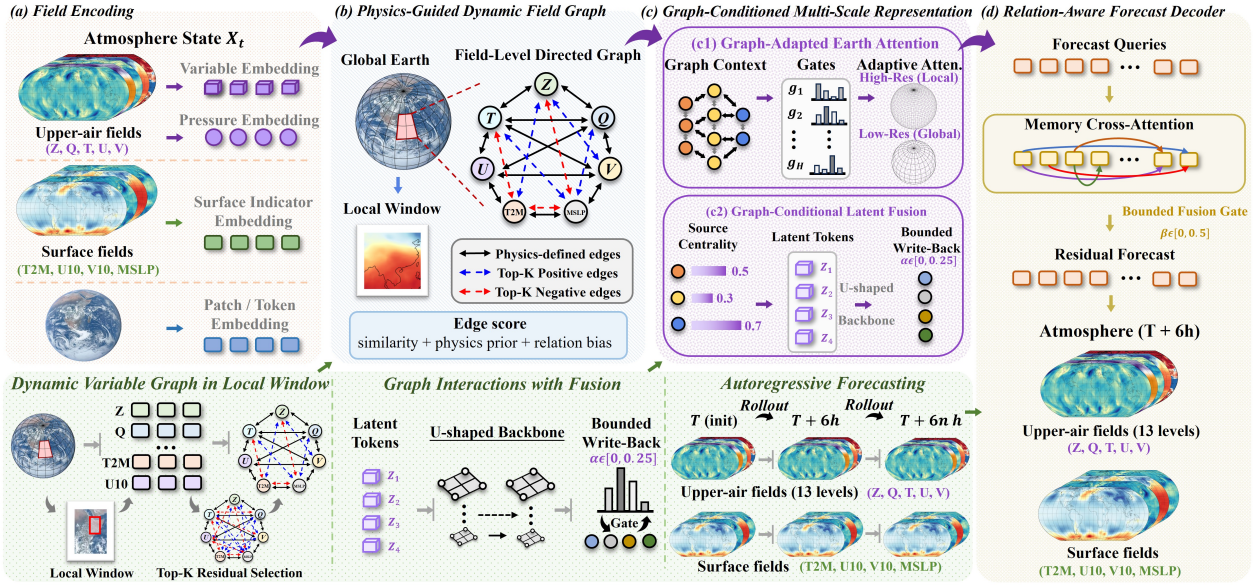


Figure 2 | The overall technical flowchart of VeinCast.

into neural forecasting. NeuralGCM and ClimODE introduce differentiable dynamics or physics-informed differential equations [Verma, Heinonen, and Garg \(2024\)](#), WeatherGFT embeds governing partial differential equations, and NowcastNet combines physical evolution with conditional generation [Zhang et al. \(2023\)](#). These methods primarily inject physics at the equation or evolution level.

Existing global medium-range models generally mix fields implicitly or construct graphs over spatial grids or meshes or incorporate equation-level physical guidance. Unlike these approaches, VeinCast organizes variable–pressure-level fields as graph nodes within local spatial windows and combines a physics-defined relational backbone with state-dependent residual connections. To the best of our knowledge, this is the first such framework for global medium-range weather forecasting.

Method

Figure. 2 provides an overview of VeinCast. It encodes surface and upper-air fields, constructs a Physics-Guided Dynamic Field Graph to adapt Earth attention and guide Graph-Conditioned Latent Fusion, and finally recovers the standard 69-field forecast through a Relation-Aware Forecast Decoder.

Problem Formulation

Given the normalized global atmospheric state $\mathbf{X}_t \in \mathbb{R}^{B \times M \times H \times W}$ at time t , VeinCast learns a state-transition model with a base interval of $\delta = 6$ h:

$$\hat{\mathbf{X}}_{t+\delta} = \mathcal{F}_\theta \left(\mathbf{X}_t, \mathbf{S}, \mathbf{P}_t, \mathbf{e}_{t+\delta}^{\text{cal}} \right), \quad (1)$$

where $\mathbf{S}$ denotes static geographic features, $\mathbf{P}_t$ is a field-availability mask used to exclude invalid source fields, and $\mathbf{e}_{t+\delta}^{\text{cal}}$ represents calendar conditions at the target valid time. We use $M = 69$ meteorological fields, consisting of 65 upper-air fields from five variables at 13 pressure levels and four surface fields. Long-range forecasts are obtained by recursively applying the 6-hour transition model. VeinCast

combines field encoding with a physics-guided dynamic field graph and graph-conditioned latent fusion, followed by a relation-aware decoder for the registered 69 forecast fields.

Field Encoding

Surface and upper-air fields have different vertical semantics and are therefore processed by two lightweight stems with independent parameters. Each scalar field is projected into spatial patch tokens and augmented with field metadata:

$$\mathbf{e}_m = \mathbf{E}_{\text{var}}(\nu_m) + \mathbf{E}_{\text{pres}}(\tilde{p}_m, s_m) + \mathbf{E}_{\text{surf}}(s_m), \quad (2)$$

where ν_m is the variable identity, $s_m \in \{0, 1\}$ is the surface indicator, and $\tilde{p}_m = \log(p_m/1000)$ for upper-air fields while $\tilde{p}_m = 0$ for surface fields. The land–sea mask, orography, and spherical positional features are projected separately and added to all field representations. The two stems are then merged in the canonical 69-field order and processed by a shared multi-scale Earth-aware encoder.

Physics-Guided Dynamic Field Graph

Regional Descriptors

VeinCast constructs a directed graph within each local spatial window, where nodes represent meteorological fields rather than latitude–longitude locations. Given $\mathbf{X} \in \mathbb{R}^{B \times M \times h \times w \times d}$, the grid is partitioned into R windows, and the T spatial tokens of each field are averaged into a regional descriptor:

$$\mathbf{D}_{b,r,m} = \frac{1}{T} \sum_{\tau=1}^T \mathbf{X}_{b,m,r,\tau}^{\text{win}}. \quad (3)$$

Physical Relations

For target field i and source field j , the relation score combines state-dependent similarity, a relation-type bias, and a physics-derived prior:

$$u_{b,r,i,j} = \frac{(\mathbf{W}_q \mathbf{D}_{b,r,i})^\top (\mathbf{W}_k \mathbf{D}_{b,r,j})}{\sqrt{d}} + e_{\bar{\rho}_{ij}} + \lambda P_{ij}^{\text{phy}}. \quad (4)$$

Here, $P_{ij}^{\text{phy}} \in \{0, 1\}$ indicates whether a predefined rule permits information flow from j to i , and $\bar{\rho}_{ij}$ is its relation type. The prior contains self, within-variable vertical, wind–geopotential, temperature–moisture, moisture–transport, and surface–atmosphere relations. Pairs without a predefined relation share a learnable generic residual type. Thus, the rules specify a stable relational support while the edge strengths remain conditioned on the local atmospheric state.

Dynamic Residual Edges

For each sample, window, and target field, the source set $\Omega_{b,r,i}$ contains all physics-defined edges and the K highest-scoring residual edges among the remaining candidates; sources masked by $\mathbf{P}_t$ are excluded. The normalized relation weights and relation-specific context are

$$\mathbf{A}_{b,r,i,j} = \text{Softmax}_{j \in \Omega_{b,r,i}} (u_{b,r,i,j}), \quad (5)$$

$$\mathbf{C}_{b,r,i} = \sum_{\ell=1}^{N_\rho} \sum_{j \in \Omega_{b,r,i}} \mathbf{A}_{b,r,i,j} \mathbf{1}[\bar{\rho}_{ij} = \ell] \mathbf{W}_v^\ell \mathbf{D}_{b,r,j}, \quad (6)$$

where each relation type has an independent value projection W_v^ℓ . The physics-defined edges form the relational backbone, whereas the top- K residual edges capture state-dependent dependencies not covered by the rule set.

Graph-Adapted Earth Attention

The graph context is broadcast to the tokens of its spatial window and combined with field and static features to generate an adapted key and a per-head gate:

$$K^{\text{ada}} = W_{\text{ada}} f_{\text{ada}}([X, C, S]), \quad (7)$$

$$g^{\text{key}} = \sigma(f_{\text{gate}}([X, C, S])), \quad (8)$$

$$K = g^{\text{key}} \odot K^{\text{base}} + (1 - g^{\text{key}}) \odot K^{\text{ada}}. \quad (9)$$

The dynamic graph therefore modulates Earth-window spatial attention without replacing its native spatial representation.

Graph-Conditioned Latent Fusion

Graph Context and Source-Node Centrality

At both spatial resolutions, the encoder alternates dynamic graph modules with Earth-window attention. The final low-resolution graph is also reused to condition multivariate fusion. Since $A_{b,r,i,j}$ measures information flow from source j to target i , the source centrality of field j and its latent-specific bias are defined as

$$c_{b,r,j} = \frac{1}{M} \sum_{i=1}^M A_{b,r,i,j}, \quad (10)$$

$$b_{b,r,n,j}^{\text{cent}} = \left[f_{\text{cent}}(\log(\max(c_{b,r,j}, \epsilon))) \right]_n, \quad (11)$$

where n indexes a fusion latent. The bias is broadcast to the corresponding spatial window. After enriching low-resolution fields with graph context, the fusion latents aggregate the available fields through

$$\tilde{X}^l = W_x[X^l, C^l], \quad (12)$$

$$Z^1 = Z^0 + \text{MHA}(Z^0, \tilde{X}^l, \tilde{X}^l; b^{\text{cent}}, P_t). \quad (13)$$

The graph context modifies field values, while source-node centrality biases field-to-latent aggregation toward locally influential source fields.

Latent U-Backbone

The fusion latents are initialized by learnable embeddings, static geography, and calendar conditions. They are processed by a Latent U-Backbone whose blocks alternate Earth-aware spatial attention within each latent slot and self-attention across latent slots at each location. Downsampling, a high-capacity bottleneck, and skip-connected upsampling capture multi-scale spatial dependencies while preserving interactions among latent semantics.

Bounded Fusion Feedback

The fused memory is written back to the low-resolution field representations by cross-attention. Let $\Delta X^l = \text{MHA}(X^l, Z, Z)$. A field- and location-dependent gate is computed from the field representation, fusion response, and field metadata:

$$\alpha = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min}) \sigma \left(f_{\alpha} \left([X^l, \Delta X^l, e] \right) \right), \quad (14)$$

$$\bar{X}^l = X^l + \alpha \odot \Delta X^l, \quad (15)$$

where $(\alpha_{\min}, \alpha_{\max}) = (0, 0.25)$. The bounded feedback preserves field-specific information while allowing each field to recover context from the shared multivariate memory.

Relation-Aware Forecast Decoder

For each of the registered 69 output fields, the decoder constructs a forecast query from its variable identity, pressure metadata and surface indicator:

$$q_i = E_{\text{var}}(v_i) + E_{\text{pres}}(\tilde{p}_i, s_i) + E_{\text{surf}}(s_i). \quad (16)$$

The query is combined with the local mean of available field features and performs relation-aware cross-attention over the low-resolution field memory:

$$Q_{\text{field}}^l = \text{MHA} \left(q + \bar{x}, \bar{X}^l, \bar{X}^l; \eta, P_t \right), \quad (17)$$

where η is a fixed bias derived from the same registered field relations, favoring the identical field, vertically related levels, and physically related variables. The field-derived query then reads the fused latent memory and combines the two sources through a bounded gate:

$$Q_{\text{fusion}}^l = \text{MHA} \left(Q_{\text{field}}^l, Z, Z \right), \quad (18)$$

$$\beta = \beta_{\min} + (\beta_{\max} - \beta_{\min}) \cdot \sigma \left(f_{\beta} \left([Q_{\text{field}}^l, Q_{\text{fusion}}^l, q] \right) \right), \quad (19)$$

$$Q^l = (1 - \beta) \odot Q_{\text{field}}^l + \beta \odot Q_{\text{fusion}}^l, \quad (20)$$

where $(\beta_{\min}, \beta_{\max}) = (0, 0.5)$. Low-resolution queries are upsampled and fused with high-resolution field-attention outputs, followed by Earth-aware spatial blocks and patch recovery. For the standard 69-field forecast, the model predicts a residual update in normalized space:

$$\widehat{X}_{t+\delta} = \text{SC}_L(X_t + \gamma R_{\theta}), \quad (21)$$

$$\text{SC}_L(x) = L \tanh(x/L), \quad (22)$$

where R_{θ} is the decoded residual and γ controls its magnitude.

Training Objective

VeinCast is trained only with a latitude-weighted Huber loss in normalized space. Let $e_{b,m,h,w}^{(s)} = \widehat{X}_{b,m,h,w}^{(s)} - X_{b,m,h,w}^{(s)}$ denote the error at rollout step s , and let $\omega_h = \cos(\phi_h) / (H^{-1} \sum_{h'=1}^H \cos(\phi_{h'}))$ be the normalized weight at latitude ϕ_h . The loss for one forecast step is

$$\mathcal{L}^{(s)} = \frac{\sum_{b,m,h,w} \omega_h \mathcal{H}_{\kappa} \left(e_{b,m,h,w}^{(s)} \right)}{\sum_{b,m,h,w} \omega_h}, \quad (23)$$

Table 1 | Three-stage training configuration. Steps, BS/GPU, Accum., and TF denote the rollout steps, per-GPU batch size, gradient-accumulation steps, and teacher-forcing probability, respectively.

Stage	Steps	Epochs	BS/GPU	Accum.	LR	TF
1	1	150	6	1	1×10^{-4}	0
2	2	10	6	4	1×10^{-5}	0.25
3	4	10	6	4	1×10^{-5}	0.25

$$\mathcal{H}_\kappa(e) = \begin{cases} \frac{1}{2}e^2, & |e| \leq \kappa, \\ \kappa \left(|e| - \frac{1}{2}\kappa \right), & |e| > \kappa, \end{cases} \quad (24)$$

$$\mathcal{L}_{\text{train}} = \frac{1}{S} \sum_{s=1}^S \mathcal{L}^{(s)}. \quad (25)$$

Here, S is the number of supervised rollout steps. No auxiliary reconstruction term or physical regularizer is used in the reported experiments.

Experiments

Dataset and Implementation Details

Dataset. We conduct experiments on the ERA5 reanalysis dataset at 6-hour intervals and a spatial resolution of 1.5° (121×240 grids) provided via WeatherBench2 [Rasp et al. \(2024\)](#); [Wedi et al. \(2015\)](#). Data from 1979–2017, 2018, and 2020 are used for training, validation, and testing, respectively. Each atmospheric state contains 69 input and output fields: geopotential (z), temperature (t), specific humidity (q), and zonal and meridional winds (u and v) at 13 pressure levels, ($\{300, 400, 500, 600, 700, 850, 925, 1000\}$ hPa), together with four surface variables: T2M, U10, V10, and mean sea-level pressure (MSL). Per-field normalization statistics are computed only from the training set. VeinCast and baseline models predict a 6-hour state transition and generate longer-range forecasts autoregressively.

Implementation Details. VeinCast and baseline models are trained on 16 NVIDIA A100 GPUs. VeinCast using AdamW with a weight decay of 10^{-3} and a linear warm-up followed by cosine learning-rate decay. We optimize a latitude-weighted Huber loss with threshold $\kappa = 2.0$. Training follows a three-stage rollout curriculum, with each stage initialized from the best checkpoint of the preceding stage. The three-stage training configurations are shown in Table 1. All baseline models are trained under a unified experimental framework, and their hyperparameters are systematically optimized to obtain the best-performing checkpoints for a fair comparison with VeinCast.

Model Performance and Analysis

Table 2 compares VeinCast with five representative global weather forecasting models, including FengWu, GraphCast, Pangu-Weather, ARROW [Tian et al. \(2026\)](#), and FuXi. We report results for three surface variables (T2M, U10, and V10) and three upper-air fields (Q500, T850, and Z700) at lead times from 3 to 14 days. Performance is evaluated using latitude-weighted root mean square error (RMSE) and anomaly correlation coefficient (ACC), where lower RMSE and higher ACC indicate better forecast accuracy.

VeinCast achieves the lowest RMSE for most variable–lead-time combinations while maintaining consistently high ACC at extended forecast horizons. Across all five lead times, it delivers the best

Table 2 | RMSE and ACC of VeinCast and baseline models on selected variables. The best results are highlighted in **bold**, and the second-best are underlined.

Methods		FengWu		GraphCast		Pangu-Weather		ARROW		FuXi		VeinCast	
Variable	Lead Time	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑
T2M	3-day	3.29	0.985	3.83	0.968	<u>2.20</u>	0.990	4.06	0.985	5.08	0.903	2.17	<u>0.990</u>
	5-day	4.44	0.976	4.82	0.949	<u>2.94</u>	<u>0.982</u>	6.75	0.967	5.80	0.895	2.75	0.984
	7-day	5.28	0.969	5.59	0.932	<u>3.69</u>	<u>0.974</u>	9.28	0.943	6.49	0.886	3.27	0.977
	10-day	6.18	0.961	6.40	0.912	<u>4.40</u>	<u>0.965</u>	12.87	0.888	7.37	0.874	3.89	0.968
	14-day	6.91	0.954	6.98	0.896	<u>4.88</u>	<u>0.958</u>	18.36	0.718	8.33	0.852	4.47	0.959
U10	3-day	2.64	0.878	2.96	0.861	2.85	0.860	<u>2.34</u>	0.910	2.93	0.822	2.31	<u>0.908</u>
	5-day	3.65	0.765	4.42	0.702	3.66	0.767	<u>3.57</u>	<u>0.794</u>	3.88	0.728	3.29	0.816
	7-day	4.32	0.668	5.47	0.561	<u>4.27</u>	<u>0.682</u>	4.58	0.664	4.61	0.634	4.07	0.720
	10-day	4.85	0.578	6.26	0.442	<u>4.81</u>	<u>0.598</u>	5.57	0.515	5.24	0.540	4.73	0.622
	14-day	5.15	0.523	6.66	0.382	<u>5.13</u>	<u>0.548</u>	6.51	0.383	5.61	0.476	5.10	0.554
V10	3-day	2.70	0.809	2.88	0.796	2.90	0.779	<u>2.38</u>	<u>0.855</u>	2.87	0.772	2.36	0.862
	5-day	3.75	0.625	4.34	0.557	3.76	0.622	<u>3.63</u>	<u>0.679</u>	3.86	0.614	3.42	0.698
	7-day	4.49	0.461	5.35	0.345	<u>4.43</u>	<u>0.473</u>	<u>4.66</u>	<u>0.481</u>	4.65	0.456	4.26	0.530
	10-day	5.05	0.323	6.12	0.163	<u>5.01</u>	<u>0.331</u>	5.61	0.281	5.28	0.305	4.95	0.366
	14-day	5.31	0.256	6.40	0.089	5.28	<u>0.259</u>	6.39	0.161	5.63	0.222	<u>5.31</u>	0.271
Q500	3-day	0.68	0.855	0.69	0.867	0.68	0.856	<u>0.64</u>	<u>0.874</u>	0.73	0.805	0.60	0.885
	5-day	0.85	0.775	0.94	0.758	0.86	0.762	<u>0.83</u>	<u>0.786</u>	0.88	0.738	0.75	0.824
	7-day	<u>0.96</u>	<u>0.712</u>	1.13	0.657	0.97	0.691	0.99	0.700	0.99	0.678	0.86	0.768
	10-day	<u>1.05</u>	<u>0.650</u>	1.29	0.554	1.05	0.628	1.18	0.589	1.10	0.611	0.97	0.701
	14-day	<u>1.12</u>	<u>0.603</u>	1.40	0.473	<u>1.10</u>	0.588	1.43	0.454	1.19	0.551	1.06	0.642
T850	3-day	2.32	0.980	2.34	0.981	<u>2.10</u>	<u>0.986</u>	2.17	0.985	3.86	0.903	1.85	0.987
	5-day	3.28	0.960	3.66	0.954	<u>3.05</u>	<u>0.970</u>	3.11	0.967	4.67	0.889	2.56	0.975
	7-day	4.07	0.940	4.68	0.926	<u>3.89</u>	<u>0.952</u>	4.08	0.941	5.37	0.871	3.30	0.958
	10-day	4.84	0.917	5.63	0.894	<u>4.72</u>	<u>0.932</u>	5.42	0.893	6.15	0.848	4.07	0.935
	14-day	5.42	0.899	6.21	0.872	<u>5.32</u>	<u>0.916</u>	7.52	0.807	6.76	0.826	4.59	0.918
Z700	3-day	310.70	0.986	284.32	0.987	265.80	0.989	249.80	0.992	514.46	0.910	209.80	0.993
	5-day	502.52	0.964	515.32	0.958	<u>436.67</u>	0.969	438.87	<u>0.974</u>	685.44	0.893	388.12	0.976
	7-day	660.34	0.937	732.71	0.916	<u>580.83</u>	<u>0.945</u>	633.77	0.943	842.80	0.867	559.86	0.950
	10-day	809.71	0.906	966.41	0.865	723.39	<u>0.914</u>	872.87	0.890	1007.41	0.833	<u>733.77</u>	0.914
	14-day	905.62	0.885	1133.90	0.834	812.65	0.891	1127.74	0.823	1123.30	0.804	<u>847.07</u>	<u>0.885</u>

overall performance for T2M, Q500, and T850, with particularly clear advantages for temperature and moisture prediction. It also performs strongly on near-surface winds, achieving the highest V10 ACC at every lead time and the highest U10 ACC from 5 to 14 days. These results highlight the benefit of explicitly modeling cross-field relationships and their joint temporal evolution. Moreover, the performance gains remain stable as the forecast horizon increases, indicating improved robustness against error accumulation during recursive forecasting. Although Pangu-Weather retains a modest advantage in Z700 RMSE at 10 and 14 days, VeinCast provides stronger and more consistent performance across heterogeneous surface and upper-air variables.

Fig. 3 provides a rank-based summary of VeinCast’s performance on normalized RMSE and ACC across all forecast lead times and variables. VeinCast consistently achieves the best or second-best rank for most variables, with especially strong dominance at 120 h and 168 h, indicating superior and stable performance across the full variable set.

Ablation Study

We conduct ablation studies to examine the contributions of the dynamic field graph, graph-conditioned latent fusion, adaptive fusion gates, and model capacity. Table 3 summarizes the performance averaged over all evaluated lead times, together with the long-range results at lead times of at least 168 hours.

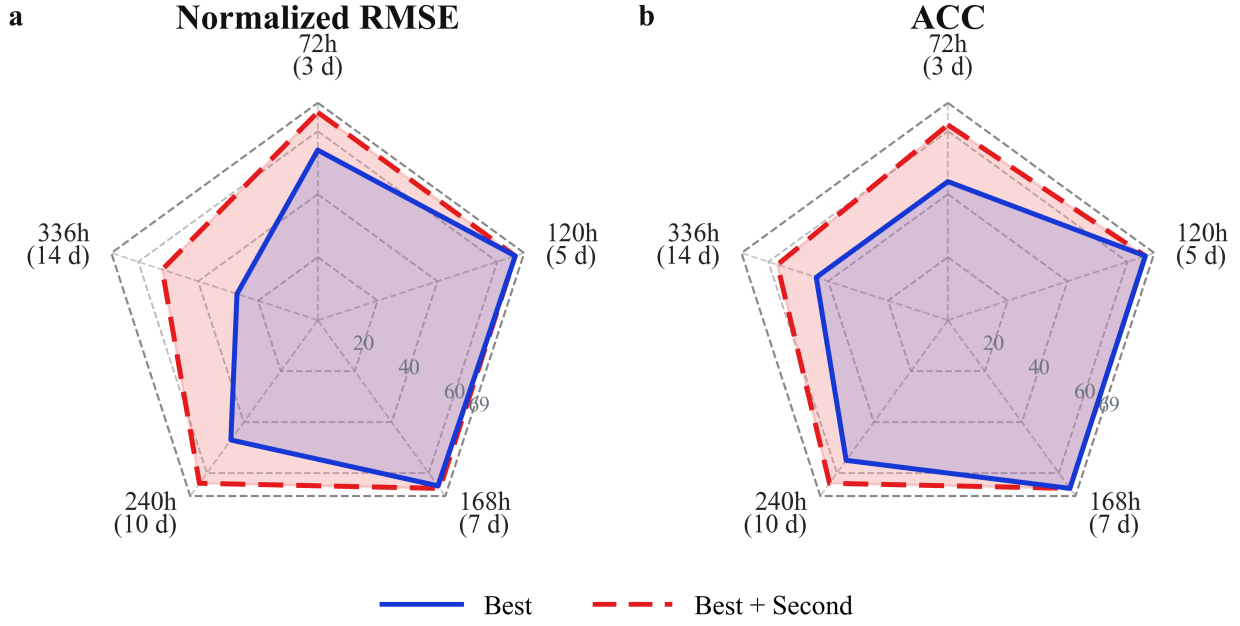


Figure 3 | Rank statistics of VeinCast on normalized RMSE and ACC across different forecast lead times. Each radar chart summarizes one evaluation metric, with spokes denoting forecast lead times. “Best” denotes the number of variables for which VeinCast ranked first among all compared methods, and “Best + Second” denotes the cumulative number of variables for which VeinCast ranked first or second.

The complete VeinCast consistently provides strong overall and long-range forecasting performance.

Dynamic field graph. Removing the dynamic field graph leads to a pronounced degradation in both normalized RMSE and ACC, with the performance gap becoming more evident at longer lead times. Retaining only the predefined physical relations recovers a substantial portion of this loss, demonstrating that the physical relational scaffold provides an effective structural prior. Nevertheless, the remaining gap to the complete model indicates that state-dependent residual edges complement the fixed physical graph by capturing cross-field interactions that vary with the local atmospheric state.

Graph-conditioned fusion. Disabling graph conditioning in the latent fusion module also causes a clear performance drop. In this variant, the dynamic graph remains available to the encoder, but its contextual representation and centrality information are not used to guide cross-field aggregation. The resulting degradation shows that relational information is most effective when it directly conditions the fusion and feedback processes, rather than remaining isolated within the graph-enhanced encoder.

Adaptive fusion gates. Replacing the adaptive α and β gates with fixed coefficients produces performance close to that of the complete model. This relatively small difference suggests that adaptive gating is not the primary source of the overall improvement. Instead, it serves as a fine-grained mechanism for regulating field-specific feedback and the use of fusion memory under different atmospheric states and forecast lead times.

Model capacity. The capacity-controlled Wide Baseline increases the encoder width while removing both the dynamic graph and the latent fusion pathway, resulting in a parameter count comparable to that of VeinCast. Despite this increased capacity, it remains substantially inferior, particularly in terms of long-range ACC. This result confirms that the improvements of VeinCast arise from its relational

Table 3 | Ablation study on key components of VeinCast. N-RMSE denotes normalized RMSE; GC denotes graph-conditioned.

Methods	Average		≥ 168 h	
	N-RMSE↓	ACC↑	N-RMSE↓	ACC↑
VeinCast	0.59	0.807	0.81	0.704
w/o Dyn. Graph	0.69	0.778	0.96	0.668
Static Phys. Graph	0.64	0.787	0.88	0.681
w/o GC Fusion	0.69	0.779	0.95	0.672
Fixed α/β	0.60	0.802	0.83	0.698
Wide Baseline	0.69	0.754	0.94	0.638

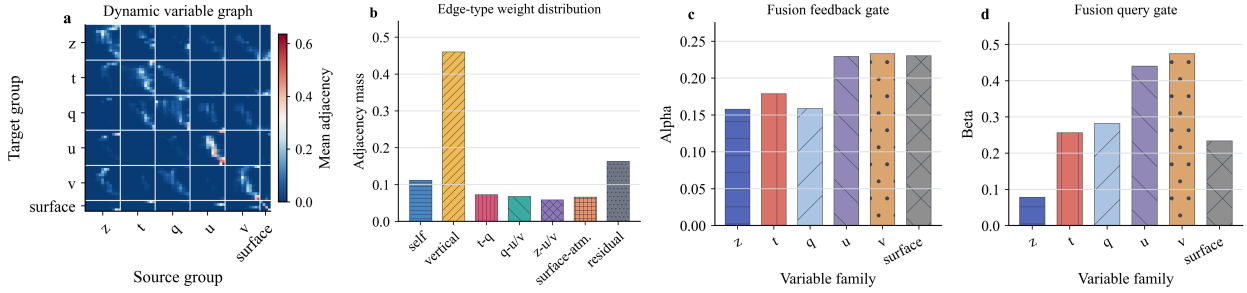


Figure 4 | Analysis of the learned graph structure and fusion gates. **(a)** Learned adjacency matrix grouped by variable family, where rows and columns denote target and source groups, respectively. **(b)** Adjacency mass assigned to different edge types. **(c)** Fusion feedback gate α for different variable families. **(d)** Fusion query gate β for different variable families.

modeling and graph-conditioned fusion mechanisms rather than from a simple increase in model size.

Progressive rollout training. Table 4 shows that the rollout fine-tuning in Stages 2 and 3 consistently improves forecasting performance over the one-step model. By progressively training the model on multi-step predictions, these stages better align the training and inference processes and mitigate error accumulation during recursive forecasting. The results therefore confirm the effectiveness of progressive rollout fine-tuning, particularly for improving the stability of long-range forecasts.

Analysis of Learned Graph Structure and Fusion Gates

We analyze the learned adjacency matrix, edge-type weights, and fusion gates to characterize how the proposed modules organize cross-variable information. The results are shown in Fig. 4.

Learned field graph. The adjacency matrix exhibits clear block-wise structures across variable families. Strong weights concentrate within the same family and between vertically related atmospheric variables, whereas unrelated variable pairs receive weaker connections. Non-zero interactions between temperature and moisture, as well as between wind and pressure-related variables, further show that the graph captures structured cross-variable dependencies rather than uniformly mixing all fields.

Edge-type weight distribution. Vertical edges account for the largest proportion of the adjacency mass, confirming that cross-level atmospheric coupling is the dominant relation in the learned graph. The learned residual receives the second-largest contribution, showing that the model supplements the predefined relations with data-driven connections. Self edges and the remaining physically defined relations retain non-zero weights, indicating that both prior structure and learned interactions

Table 4 | Ablation study on VeinCast training stages. N-RMSE denotes normalized RMSE.

Methods	Average		≥ 168 h	
	N-RMSE↓	ACC↑	N-RMSE↓	ACC↑
VeinCast (Stage 1)	0.64	0.785	0.86	0.678
VeinCast (Stage 2)	0.63	0.794	0.84	0.688
VeinCast (Stage 3)	0.59	0.807	0.81	0.704

contribute to graph construction.

Fusion feedback gate. The feedback gate α assigns larger values to the horizontal wind and surface variables than to geopotential height, temperature, and moisture. This variable-dependent pattern shows that the feedback pathway modulates different atmospheric fields unequally, with stronger modulation for wind and surface-related predictions.

Fusion query gate. The query gate β exhibits stronger variable selectivity than α . The largest gate values occur for the horizontal wind variables, whereas geopotential height receives the smallest value. The remaining variables lie between these two extremes. This ordering indicates that the query pathway adjusts the contribution of fused information according to the target variable family.

Overall, the learned graph combines structured atmospheric relations with complementary data-driven connections, while the two fusion gates apply distinct variable-dependent modulation. These results show that the proposed modules perform selective information integration across atmospheric variables.

Conclusion

This paper introduced VeinCast to model structured and state-dependent interactions among meteorological fields in global medium-range weather forecasting. The Physics-Guided Dynamic Field Graph combines predefined relations with state-dependent residual edges, while its context adapts Earth-window attention and guides field-to-latent aggregation through Graph-Conditioned Latent Fusion. Experiments show competitive accuracy across surface and upper-air variables, particularly at extended lead times. Ablations and graph analyses demonstrate that the physical scaffold, residual connections, and graph-conditioned aggregation are complementary, and that the gains do not arise from model capacity alone. These findings support relational-level physical priors as a middle ground between implicit data-driven mixing and rigid equation-level constraints. The current evaluation is limited to deterministic forecasts at 1.5° resolution over a fixed 69-field set. Future work will consider higher-resolution and probabilistic forecasting.

References

- Al-Yahyai, S.; Charabi, Y.; and Gastli, A. 2010. Review of the use of numerical weather prediction (NWP) models for wind energy assessment. *Renewable and Sustainable Energy Reviews*, 14(9): 3192–3198.
- Allen, A.; Markou, S.; Tebbutt, W.; Requeima, J.; Bruinsma, W. P.; Andersson, T. R.; Herzog, M.; Lane, N. D.; Chantry, M.; Hosking, J. S.; et al. 2025. End-to-end data-driven weather prediction. *Nature*, 641(8065): 1172–1179.

- Bauer, P.; Thorpe, A.; and Brunet, G. 2015. The quiet revolution of numerical weather prediction. *Nature*, 525(7567): 47–55.
- Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; and Tian, Q. 2023. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619: 533 – 538.
- Bodnar, C.; Bruinsma, W. P.; Lucic, A.; Stanley, M.; Allen, A.; Brandstetter, J.; Garvan, P.; Riechert, M.; Weyn, J. A.; Dong, H.; et al. 2025. A foundation model for the Earth system. *Nature*, 641(8065): 1180–1187.
- Chen, H.; Han, T.; Zhang, J.; Guo, S.; and Bai, L. 2026a. STCast: Adaptive Boundary Alignment for Global and Regional Weather Forecasting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*.
- Chen, H.; Han, T.; ZHANG, J.; Guo, S.; Ling, F.; and BAI, L. 2026b. EMFormer: Efficient Multi-Scale Transformer for Accumulative Context Weather Forecasting. In *Forty-third International Conference on Machine Learning*.
- Chen, K.; Han, T.; Gong, J.; Bai, L.; Ling, F.; Luo, J.-J.; Chen, X.; Ma, L.; Zhang, T.; Su, R.; et al. 2023a. Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. *arXiv preprint arXiv:2304.02948*.
- Chen, K.; Han, T.; Ling, F.; Gong, J.; Bai, L.; Wang, X.; Luo, J. J.; Fei, B.; Zhang, W.; and Chen, X. 2025. The operational medium-range deterministic weather forecasting can be extended beyond a 10-day lead time. *Communications Earth & Environment*, 6(1): 1–10.
- Chen, L.; Zhong, X.; and Zhang, H., Yuan Li. 2023. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj climate and atmospheric science*, 6(1).
- Chen, S.; Long, G.; Shen, T.; and Jiang, J. 2023b. Prompt federated learning for weather forecasting: Toward foundation models on meteorological data. *arXiv preprint arXiv:2301.09152*.
- Gan, L.; Man, X.; Zhang, C.; and Shao, J. 2025. EWMoE: An effective model for global weather forecasting with mixture-of-experts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 210–218.
- Gao, Y.; Wu, H.; Shu, R.; Dong, H.; Xu, F.; Chen, R. R.; Yan, Y.; Wen, Q.; Hu, X.; Wang, K.; et al. 2025. OneForecast: A Universal Framework for Global and Regional Weather Forecasting. In *International Conference on Machine Learning*, 18658–18697. PMLR.
- Keisler, R. 2022. Forecasting Global Weather with Graph Neural Networks. *arXiv:2202.07575*.
- Kochkov, D.; Yuval, J.; Langmore, I.; Norgaard, P.; Smith, J.; Mooers, G.; Klöwer, M.; Lottes, J.; Rasp, S.; Düben, P.; et al. 2024. Neural general circulation models for weather and climate. *Nature*, 632(8027): 1060–1066.
- Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W.; et al. 2023. Learning skillful medium-range global weather forecasting. *Science*, 382(6677): 1416–1421.
- Maraun, D.; Wetterhall, F.; Ireson, A. M.; Chandler, R. E.; Kendon, E. J.; Widmann, M.; Brienien, S.; Rust, H. W.; Sauter, T.; and Themel, M. 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48(3).

- Molteni, F.; Buizza, R.; Palmer, T. N.; and Petroliagis, T. 1996. The ECMWF Ensemble Prediction System: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122(529).
- Nguyen, T.; Brandstetter, J.; Kapoor, A.; Gupta, J. K.; and Grover, A. 2023. Climax: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*.
- Nguyen, T.; Shah, R.; Bansal, H.; Arcomano, T.; Maulik, R.; Kotamarthi, R.; Foster, I.; Madireddy, S.; and Grover, A. 2024. Scaling transformer neural networks for skillful and reliable medium-range weather forecasting. In Globerson, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J.; and Zhang, C., eds., *Advances in Neural Information Processing Systems*, volume 37, 68740–68771. Curran Associates, Inc.
- Oskarsson, J.; Landelius, T.; Deisenroth, M. P.; and Lindsten, F. 2024. Probabilistic Weather Forecasting with Hierarchical Graph Neural Networks. In Globerson, A.; Mackey, L.; Belgrave, D.; Fan, A.; Paquet, U.; Tomczak, J.; and Zhang, C., eds., *Advances in Neural Information Processing Systems*, volume 37, 41577–41648. Curran Associates, Inc.
- Pathak, J.; Subramanian, S.; Harrington, P.; Raja, S.; Chattopadhyay, A.; Mardani, M.; Kurth, T.; Hall, D.; Li, Z.; Azizzadenesheli, K.; et al. 2022. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*.
- Price, I.; Sanchezgonzalez, A.; Alet, F.; Andersson, T. R.; Elkadi, A.; Masters, D.; Ewalds, T.; Stott, J.; Mohamed, S.; and Battaglia, P. 2025. Probabilistic weather forecasting with machine learning. *Nature*.
- Rasp, S.; Hoyer, S.; Meroze, A.; Langmore, I.; Battaglia, P.; Russell, T.; Sanchez-Gonzalez, A.; Yang, V.; Carver, R.; Agrawal, S.; Chantry, M.; Ben Bouallegue, Z.; Dueben, P.; Bromberg, C.; Sisk, J.; Barrington, L.; Bell, A.; and Sha, F. 2024. WeatherBench 2: A Benchmark for the Next Generation of Data-Driven Global Weather Models. *Journal of Advances in Modeling Earth Systems*, 16(6): e2023MS004019. E2023MS004019 2023MS004019.
- Ren, X.; Li, X.; Ren, K.; Song, J.; Xu, Z.; Deng, K.; and Wang, X. 2021. Deep learning-based weather prediction: a survey. *Big Data Research*, 23: 100178.
- Tian, J.; Ding, Y.; Xu, R.; Miao, H.; Guo, C.; and Yang, B. 2026. ARROW: An Adaptive Rollout and Routing Method for Global Weather Forecasting. In *The Fourteenth International Conference on Learning Representations*.
- Verma, Y.; Heinonen, M.; and Garg, V. 2024. ClimODE: Climate and Weather Forecasting with Physics-informed Neural ODEs. *arXiv:2404.10024*.
- Wedi, N.; Bauer, P.; Denoninck, W.; Diamantakis, M.; Hamrud, M.; Kuhnlein, C.; Malardel, S.; Mogensen, K.; Mozdzyński, G.; and Smolarkiewicz, P. 2015. The modelling infrastructure of the Integrated Forecasting System: Recent advances and future challenges.
- Xu, W.; Ling, F.; Zhang, W.; Han, T.; Chen, H.; Ouyang, W.; and Bai, L. 2025. Generalizing Weather Forecast to Fine-grained Temporal Scales via Physics-AI Hybrid Modeling. *arXiv:2405.13796*.
- Zhang, Y.; Long, M.; Chen, K.; Xing, L.; Jin, R.; Jordan, M. I.; and Wang, J. 2023. Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*.
- Zhao, T.; Wang, S.; Ouyang, C.; Chen, M.; Liu, C.; Zhang, J.; Yu, L.; Wang, F.; Xie, Y.; Li, J.; Wang, F.; Grunwald, S.; Wong, B. M.; Zhang, F.; Qian, Z.; Xu, Y.; Yu, C.; Han, W.; Sun, T.; Shao, Z.; Qian, T.

Chen, Z.; Zeng, J.; Zhang, H.; Letu, H.; Zhang, B.; Wang, L.; Luo, L.; Shi, C.; Su, H.; Zhang, H.; Yin, S.; Huang, N.; Zhao, W.; Li, N.; Zheng, C.; Zhou, Y.; Huang, C.; Feng, D.; Xu, Q.; Wu, Y.; Hong, D.; Wang, Z.; Lin, Y.; Zhang, T.; Kumar, P.; Plaza, A.; Chanussot, J.; Zhang, J.; Shi, J.; and Wang, L. 2024. Artificial intelligence for geoscience: Progress, challenges, and perspectives. *The Innovation*, 5(5): 100691.

Appendix

A. Model Details for VeinCast

This section provides a complete specification of the VeinCast architecture and its autoregressive training procedure. VeinCast remains a closed-set forecaster: both its input and its supervised forecast target consist of the same registered 69 meteorological fields.

Notation and End-to-End Tensor Flow

At analysis time t , the normalized atmospheric state is $X_t \in \mathbb{R}^{B \times M \times H \times W}$, where B is the batch size, $M = 69$ is the number of registered fields, and $(H, W) = (121, 240)$ is the latitude–longitude grid. The field registry contains five upper-air variables, geopotential z , temperature t , specific humidity q , zonal wind u , and meridional wind v , at the 13 pressure levels

$$\mathcal{P} = \{50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, 1000\} \text{ hPa.} \quad (26)$$

together with the four surface variables T2M, U10, V10, and MSL. We index a field by $m \in \{1, \dots, M\}$ and associate it with a variable identity v_m , a pressure p_m , and a surface indicator $s_m \in \{0, 1\}$.

VeinCast also receives static geographic features $\mathcal{S} \in \mathbb{R}^{B \times C_s \times H \times W}$, a field-availability mask $\mathcal{P}_t \in \{0, 1\}^{B \times M}$, and calendar features at the forecast valid time. The availability mask acts on source fields throughout the graph, fusion, and decoder modules. A missing source therefore cannot transmit information to another field, although the tensor shape remains fixed.

Table 5 summarizes the main representations in the default configuration. The field dimension and the latent-slot dimension are kept explicit throughout the network; they are never folded into ordinary feature channels.

The full one-step computation can be written schematically as

$$\begin{aligned} (X^h, X^l, \mathcal{G}^l) &= \mathcal{E}(X_t, \mathcal{S}, \mathcal{P}_t), \\ Z &= \mathcal{U}_{\text{latent}} \left(\mathcal{A}_{F \rightarrow Z}(X^l, \mathcal{G}^l) \right), \\ \bar{X}^l &= \mathcal{A}_{Z \rightarrow F}(X^l, Z), \\ R_\theta &= \mathcal{D}(X^h, \bar{X}^l, Z, \mathcal{P}_t), \\ \hat{X}_{t+\delta} &= \text{SC}_{12}(X_t + 0.25R_\theta), \end{aligned} \quad (27)$$

where $\delta = 6$ h, $\mathcal{E}$ is the two-scale field encoder, $\mathcal{G}^l$ is the final low-resolution dynamic graph state, $\mathcal{U}_{\text{latent}}$ is the Latent U-Backbone, and $\mathcal{D}$ is the relation-aware decoder. Long-range forecasts are obtained by recursively applying this shared 6-hour transition operator.

Dual-Modality Field Encoding

Continuous Field Metadata

For an upper-air field, pressure is represented on a logarithmic coordinate; surface fields use a distinguished coordinate and an explicit surface flag:

$$\tilde{p}_m = \begin{cases} \log(p_m/1000), & s_m = 0, \\ 0, & s_m = 1. \end{cases} \quad (28)$$

Table 5 | Tensor shapes in the default VeinCast configuration. Multiplicity denotes the explicit field or latent dimension.

Representation	Spatial grid	Width	Multiplicity
Input state	121×240	1	69 fields
High-resolution fields	31×60	96	69 fields
Low-resolution fields	16×30	192	69 fields
Fusion latents	16×30	384	4 slots
Fusion bottleneck	8×15	768	4 slots
Decoded queries	31×60	96	69 queries
Forecast state	121×240	1	69 fields

Let $\{\omega_k\}_{k=1}^{K_p}$ be logarithmically spaced Fourier frequencies. The continuous pressure features are

$$\psi_p(\tilde{p}_m, s_m) = \left[\tilde{p}_m, s_m, \{\sin(\omega_k \tilde{p}_m), \cos(\omega_k \tilde{p}_m)\}_{k=1}^{K_p} \right]. \quad (29)$$

An MLP maps these features to the token width. We define

$$E_{\text{pres}}(\tilde{p}_m, s_m) := f_{\text{pres}}(\psi_p(\tilde{p}_m, s_m)).$$

The complete field metadata is

$$e_m = E_{\text{var}}(v_m) + E_{\text{pres}}(\tilde{p}_m, s_m) + E_{\text{surf}}(s_m). \quad (30)$$

The same parameterization is reused to describe output fields in the forecast decoder. In the reported closed-set experiments, all output queries correspond to the registered pressure levels and surface variables.

Surface and Upper-Air Stems

Surface and upper-air fields are processed by two independently parameterized stems. For either modality, each scalar field is first zero-padded to a multiple of the 4×4 patch size and projected by a strided convolution. A lightweight refinement block then applies group normalization, a depthwise 3×3 convolution, GELU activation, and a pointwise convolution with a residual connection. Longitude is padded periodically in the refinement convolution, whereas latitude is padded by boundary replication. The output of the stem for field m is

$$X_m^{h,0} = \text{LN}(\text{Stem}_{s_m}(X_{t,m}) + e_m + m_m), \quad (31)$$

where m_m is a learnable missing-field token when $P_{t,m} = 0$ and zero otherwise. The physical input values of an unavailable field are set to zero before the stem. The two modality-specific outputs are then restored to the canonical 69-field order.

Static inputs consist of the land–sea mask, orography, and spherical location features derived from latitude ϕ and longitude λ . They are independently patch-embedded and added to every field token. This produces $X^{h,0} \in \mathbb{R}^{B \times M \times 31 \times 60 \times 96}$.

Two-Scale Earth-Aware Encoder

At the high-resolution scale, two Earth-window blocks alternate non-shifted and shifted 4×8 windows with four attention heads. Before every block, a dynamic graph module is reconstructed from the current field representations and provides the graph-adapted keys described below. Patch merging

Table 6 | Physical relation registry for the 69-field configuration. The surface–atmosphere mappings are T2M with $\{t, q\}$, U10 with $\{u, z\}$, V10 with $\{v, z\}$, and MSL with $\{z, u, v, t\}$. Each valid cross-field pair is registered in both information-flow directions.

Relation type	Rule for source and target fields	Directed edges
Self	Identical variable and level or identical surface field	69
Vertical	Same upper-air variable at two distinct pressure levels	780
Wind–geopotential	z paired with u or v at nearby pressure levels	204
Temperature–moisture	t paired with q at nearby pressure levels	102
Moisture transport	q paired with u or v at nearby pressure levels	204
Surface–atmosphere	Registered surface/upper-air pair at $p \geq 700$ hPa	80

concatenates neighboring 2×2 tokens, applies layer normalization, and projects their combined width from 4×96 to 192. The resulting low-resolution grid is 16×30 . Static tokens are bilinearly aligned to this grid, linearly projected, and added again. Four additional Earth-window blocks operate at this scale with eight heads and alternating window shifts.

For each Earth-window block, the positional bias contains a relative-position term shared across windows and a latitude-window-specific term:

$$\mathbf{B}_r^{\text{earth}} = \mathbf{B}^{\text{rel}} + \mathbf{B}_{r_\phi}^{\text{lat}}, \quad (32)$$

where r_ϕ identifies the latitude band of window r . Periodic longitude padding respects global wraparound, while replicated latitude padding avoids introducing an artificial connection between the two poles.

Physics-Guided Dynamic Field Graph

Window-Level Field Descriptors

The graph is reconstructed independently for each sample and spatial window. Its nodes are meteorological fields, not geographic grid cells. Given field tokens $\mathbf{X} \in \mathbb{R}^{B \times M \times h \times w \times d}$, we first pad the spatial grid to a multiple of the window size and partition it into R windows containing T tokens. The regional descriptor of field m is

$$\mathbf{D}_{b,r,m} = \frac{1}{T} \sum_{\tau=1}^T \mathbf{x}_{b,m,r,\tau}^{\text{win}}. \quad (33)$$

The same descriptor is used to score candidate edges and to construct the relation-specific graph context. This compression avoids applying an $M \times M$ relation operator separately at every spatial token.

Exact Relation Registry

For a directed edge $j \rightarrow i$, field j is the source from which target field i reads information. The physical registry uses the rules in Table 6. Two upper-air fields are regarded as being at nearby pressure levels when

$$|\log p_i - \log p_j| < 0.45. \quad (34)$$

The relation rules define candidate support rather than a numerical evolution equation: every retained edge still receives a state-dependent learned weight.

Let $\rho_{ij} \in \{1, \dots, N_\rho\}$ denote a physical relation type and P_{ij}^{phy} indicate whether a rule permits $j \rightarrow i$. Pairs not covered by the registry are assigned a shared learned-residual type for scoring. After descriptor normalization, the edge score is

$$u_{b,r,i,j} = d^{-1/2} (W_q \text{LN}(\mathbf{D}_{b,r,i}))^\top \cdot (W_k \text{LN}(\mathbf{D}_{b,r,j})) + e_{\tilde{\rho}_{ij}} + \lambda P_{ij}^{\text{phy}}, \quad (35)$$

where $e_{\tilde{\rho}_{ij}}$ is a learned scalar bias and $\lambda = 1$ in the default configuration. The relation prior raises the initial preference for a registered physical edge without fixing its final attention weight.

Residual Edge Selection and Normalization

Source availability is applied before residual selection:

$$u_{b,r,i,j} \leftarrow -\infty \quad \text{if} \quad P_{t,b,j} = 0. \quad (36)$$

For every (b, r, i) , all available physics-defined edges are retained. Among the remaining candidates, the $K = 4$ highest-scoring edges are selected as learned residuals. With $\Omega_{b,r,i}$ denoting the union of these two sets, the normalized adjacency is

$$A_{b,r,i,j} = \frac{\exp(u_{b,r,i,j})}{\sum_{k \in \Omega_{b,r,i}} \exp(u_{b,r,i,k})}, \quad j \in \Omega_{b,r,i}, \quad (37)$$

and zero otherwise. Top- K selection is thus local to the sample, spatial window, and target field. A residual edge may change across locations or forecast states even though the physical scaffold remains fixed.

Each relation type has an independent value projection. Let $\tilde{\rho}_{b,r,i,j}$ be the final edge type, with all selected nonphysical edges sharing the learned-residual type. The field-specific graph context is

$$\mathbf{C}_{b,r,i} = \sum_{\ell=1}^{N_\rho} \sum_{j \in \Omega_{b,r,i}} A_{b,r,i,j} \mathbf{1}[\tilde{\rho}_{b,r,i,j} = \ell] W_v^\ell \mathbf{D}_{b,r,j}. \quad (38)$$

This context is broadcast to all T tokens of the corresponding window. The graph construction has complexity $O(BRM^2d)$ for a fixed windowing scheme, rather than applying the same pairwise field operation independently to every spatial token.

Graph-Adapted Earth-Window Attention

The graph communicates cross-field information to a spatial block by adapting its key representation. At each field token, graph context and static context are concatenated with the current representation:

$$\mathbf{H}^{\text{ada}} = [\mathbf{X}, \mathbf{C}, \mathbf{S}]. \quad (39)$$

Two small networks generate an adapted key and one gate per attention head:

$$\begin{aligned} \mathbf{K}^{\text{ada}} &= W_{\text{ada}} f_{\text{ada}}(\mathbf{H}^{\text{ada}}), \\ \mathbf{g}^{\text{key}} &= \sigma \left(f_{\text{gate}}(\mathbf{H}^{\text{ada}}) \right). \end{aligned} \quad (40)$$

After spatial window partitioning, the key used by each attention head is

$$\mathbf{K} = \mathbf{g}^{\text{key}} \odot \mathbf{K}^{\text{base}} + (1 - \mathbf{g}^{\text{key}}) \odot \mathbf{K}^{\text{ada}}. \quad (41)$$

The corresponding query and value remain those of the native Earth-window block. Therefore, the graph modulates within-field spatial attention without replacing its spatial representation: a gate near one preserves the base key, whereas a smaller gate increases the graph-adapted contribution.

Graph-Conditioned Latent Fusion

Conditioned Latent Initialization

At every position of the 16×30 grid, VeinCast maintains $N_z = 4$ latent slots with width $d_z = 384$. Their initialization combines learnable latent identities with static, geographic, lead-time, and calendar conditions:

$$\mathbf{Z}^0 = \mathbf{Z}^{\text{learn}} + E_{\text{static}}(\mathbf{S}^l) + E_{\text{geo}}(G_{\text{geo}}^l) + E_{\text{lead}}(\delta/24) + E_{\text{cal}}(m_{t+\delta}, h_{t+\delta}). \quad (42)$$

Here G_{geo}^l contains normalized latitude and longitude coordinates and multi-frequency sine/cosine features. Month and UTC hour are encoded cyclically. Invalid or unavailable calendar values are replaced by a learned unknown-time embedding. Although E_{lead} is continuously parameterized, every transition in the reported model uses the trained base interval $\delta = 6$ h.

Graph Context and Source Centrality

The final low-resolution graph state is reused rather than reconstructed solely for fusion. First, field tokens are enriched with their graph context:

$$\tilde{\mathbf{X}}^l = W_x \text{LN}([\mathbf{X}^l, \mathbf{C}^l]) \in \mathbb{R}^{B \times M \times 16 \times 30 \times 384}. \quad (43)$$

Second, because $A_{b,r,i,j}$ measures information flow from source j to target i , the average outgoing contribution of source j is

$$c_{b,r,j} = \frac{1}{M} \sum_{i=1}^M A_{b,r,i,j}. \quad (44)$$

After lower bounding by $\epsilon = 10^{-4}$ and applying a logarithm, a small MLP maps this scalar to one bias for each latent slot:

$$b_{b,r,n,j}^{\text{cent}} = [f_{\text{cent}}(\log(\max(c_{b,r,j}, \epsilon)))]_n. \quad (45)$$

The bias is shared across attention heads but varies with sample, location, latent slot, and source field.

At each low-resolution spatial position, fusion latents serve as queries and fields serve as keys and values. For head a , latent slot n , and source field m , the attention logit is

$$\ell_{b,y,x,a,n,m} = d_a^{-1/2} \langle W_q^a \text{LN}(\mathbf{Z}_{b,n,y,x}^0), W_k^a \tilde{\mathbf{X}}_{b,m,y,x}^l \rangle + b_{b,y,x,n,m}^{\text{cent}}. \quad (46)$$

Unavailable source fields are masked before normalization:

$$\pi_{b,y,x,a,n,m} = \text{Softmax}_{m:P_{t,b,m}=1}(\ell_{b,y,x,a,n,m}). \quad (47)$$

The multi-head output is projected and added residually to $\mathbf{Z}^0$. Graph context and source centrality therefore play complementary roles: graph context changes the field representation being read, while centrality biases the relative priority of locally influential sources.

Latent U-Backbone

The graph-conditioned latents are processed by a U-shaped hierarchy. Three blocks operate at 16×30 with width 384 and 12 heads. Patch merging then produces an 8×15 representation with width 768, followed by four bottleneck blocks with 24 heads. Bilinear patch expansion returns to 16×30 and width 384. The expanded features are concatenated with the encoder skip representation, linearly fused, and processed by two decoder blocks with 12 heads.

Every fusion block contains two forms of interaction. First, Earth-aware window attention and an MLP mix spatial tokens independently inside each latent slot. Second, multi-head self-attention and an MLP mix the four latent slots at each fixed geographic location. If $\mathcal{B}_s$ and $\mathcal{B}_z$ denote these two operations, a fusion block is

$$\mathbf{Z}' = \mathcal{B}_z(\mathcal{B}_s(\mathbf{Z})). \quad (48)$$

This factorization separates multiscale spatial propagation from interaction among latent semantics.

Bounded Fusion-to-Field Feedback

The fused latent memory is returned to the low-resolution field pathway by cross-attention at each spatial location:

$$\Delta \mathbf{X}^l = \text{MHA} (Q = \text{LN}(\mathbf{X}^l), K = \text{LN}(\mathbf{Z}), V = \text{LN}(\mathbf{Z})). \quad (49)$$

The amount written to each field and location is bounded:

$$\alpha = 0.25 \sigma \left(f_\alpha([\mathbf{X}^l, \Delta \mathbf{X}^l, \mathbf{e}]) \right), \quad (50)$$

$$\bar{\mathbf{X}}^l = \mathbf{X}^l + \alpha \odot \Delta \mathbf{X}^l. \quad (51)$$

Thus $\alpha_{b,m,y,x} \in (0, 0.25)$. The field-dependent gate allows shared multivariate information to be recovered selectively while preventing the fusion pathway from overwhelming field-specific representations.

Relation-Aware Forecast Decoder

Query Construction and Relation Prior

For each registered output field i , the decoder constructs metadata using the same variable, pressure, and surface parameterization as the encoder:

$$\mathbf{q}_i = \mathbf{E}_{\text{var}}(v_i) + \mathbf{E}_{\text{pres}}(\tilde{p}_i, s_i) + \mathbf{E}_{\text{surf}}(s_i). \quad (52)$$

A continuous embedding of the 6-hour transition interval is added at both decoder resolutions. The query is also conditioned on the local mean of the available field memory:

$$\begin{aligned} \bar{\mathbf{x}}_{b,y,x} &= \frac{\sum_{m=1}^M P_{t,b,m} \mathbf{X}_{b,m,y,x}}{\max(\sum_{m=1}^M P_{t,b,m}, 1)}, \\ h_{b,i,y,x}^0 &= \mathbf{q}_i + \bar{\mathbf{x}}_{b,y,x}. \end{aligned} \quad (53)$$

The decoder uses the same relation registry as the dynamic graph, but translates it into a fixed soft bias between query i and input field m . Let $r(i, m)$ denote their registered relation:

$$\eta_{i,m} = \begin{cases} 2.0, & r(i, m) = \text{self}, \\ 1.5 - |\log p_i - \log p_m|, & r(i, m) = \text{vertical}, \\ 0.75, & r(i, m) \in \mathcal{R}_{\text{physical}}^*, \\ -2.0, & r(i, m) = \text{none}, \end{cases} \quad (54)$$

where $\mathcal{R}_{\text{physical}}^*$ contains all registered physical relations other than self and vertical. This is a preference rather than a hard mask; an unrelated but available field can still be read if its learned content score is sufficiently informative.

Low- and High-Resolution Memory Reading

At the low-resolution scale, query i attends to the updated field memory $\bar{\mathbf{X}}^l$ at every location. For attention head a ,

$$s_{b,y,x,a,i,m}^l = \frac{\langle W_q^a \text{LN}(\mathbf{h}_{b,i,y,x}^0), W_k^a \bar{\mathbf{X}}_{b,m,y,x}^l \rangle}{\sqrt{d_a}} + \eta_{i,m}. \quad (55)$$

Unavailable fields are masked before softmax. The resulting field-derived query $\mathbf{Q}_{\text{field}}^l$ subsequently reads the four fusion latents:

$$\mathbf{Q}_{\text{fusion}}^l = \text{MHA}(\mathbf{Q}_{\text{field}}^l, \mathbf{Z}, \mathbf{Z}). \quad (56)$$

Their contributions are combined by a query- and location-dependent bounded gate,

$$\beta = 0.5 \sigma \left(f_\beta([\mathbf{Q}_{\text{field}}^l, \mathbf{Q}_{\text{fusion}}^l, \mathbf{q}]) \right), \quad (57)$$

$$\mathbf{Q}^l = (1 - \beta) \odot \mathbf{Q}_{\text{field}}^l + \beta \odot \mathbf{Q}_{\text{fusion}}^l, \quad (58)$$

where $\beta_{b,i,y,x} \in (0, 0.5)$.

The low-resolution queries are projected and bilinearly expanded to 31×60 . In parallel, an independent high-resolution query-to-field attention reads $\mathbf{X}^h$ using the same metadata, local context, source mask, and relation bias. The high-resolution field output and expanded low-resolution query are concatenated and linearly fused. Two Earth-aware query spatial blocks, one non-shifted and one shifted, then model spatial dependencies within every output field before patch recovery.

Patch Recovery and Stable Residual Update

Field patch recovery linearly projects each query token to the 16 scalar values of its 4×4 output patch. Let $\mathbf{U}_{\theta,v}$ be the raw output for variable v . It is first softly bounded and then transformed by a variable-specific affine map:

$$s_v = 1 + 0.25 \tanh(a_v), \quad b_v = 0.25 \tanh(c_v), \quad (59)$$

$$\mathbf{R}_{\theta,v} = s_v \text{SC}_8(\mathbf{U}_{\theta,v}) + b_v, \quad \text{SC}_L(x) = L \tanh(x/L). \quad (60)$$

Pressure levels of the same variable share (s_v, b_v) . For the default 69-field query set, the final forecast is the normalized-space residual update

$$\hat{\mathbf{X}}_{t+6h} = \text{SC}_{12}(\mathbf{X}_t + 0.25\mathbf{R}_\theta). \quad (61)$$

The internal SC_8 limits the decoded update, while the final SC_{12} limits the recursively propagated state. These operations are numerical stabilization mechanisms and do not impose mass, energy, or momentum conservation.

Autoregressive Training and Inference

Masked Latitude-Weighted Huber Objective

The reported model is optimized with a latitude-weighted Huber forecast loss in per-field normalized space. Let $\mu_{b,m,h,w}^{(s)}$ indicate that the target location is valid and participates in supervision. In particular, pressure levels located below the local surface pressure can be excluded. With

$$\omega_h = \frac{\max(\cos \phi_h, 0)}{H^{-1} \sum_{h'=1}^H \max(\cos \phi_{h'}, 0)}, \quad (62)$$

the loss at rollout step s is

$$\mathcal{L}^{(s)} = \frac{\sum_{b,m,h,w} \mu_{b,m,h,w}^{(s)} \omega_h \mathcal{H}_\kappa(e_{b,m,h,w}^{(s)})}{\max\left(\sum_{b,m,h,w} \mu_{b,m,h,w}^{(s)} \omega_h, 1\right)}, \quad (63)$$

where $e^{(s)} = \widehat{\mathbf{X}}^{(s)} - \mathbf{X}^{(s)}$ and

$$\mathcal{H}_\kappa(e) = \begin{cases} \frac{1}{2}e^2, & |e| \leq \kappa, \\ \kappa(|e| - \frac{1}{2}\kappa), & |e| > \kappa, \end{cases} \quad \kappa = 2. \quad (64)$$

The normalization of ω_h gives the latitude weights unit mean. No auxiliary reconstruction loss or equation-based physical regularizer is used in the reported objective. For an S -step rollout, the total objective is

$$\mathcal{L}_{\text{train}} = \frac{1}{S} \sum_{s=1}^S \mathcal{L}^{(s)}. \quad (65)$$

Progressive Rollout Curriculum

Training proceeds from a one-step transition to longer self-generated input sequences. Stage 1 supervises one 6-hour step. Stage 2 initializes model weights from the best Stage-1 checkpoint and supervises two autoregressive steps. Stage 3 initializes from the best Stage-2 checkpoint and supervises four steps. The optimizer and learning-rate schedule are restarted at each stage.

Let $\mathbf{X}^{(0)} = \mathbf{X}_t$. At rollout step s ,

$$\widehat{\mathbf{X}}^{(s)} = \mathcal{F}_\theta\left(\mathbf{X}^{(s-1)}, \mathcal{S}, \mathbf{P}^{(s-1)}, \mathbf{e}_{t+s\delta}^{\text{cal}}\right), \quad \delta = 6 \text{ h}. \quad (66)$$

Every application is conditioned on the same base transition interval, while the target-valid calendar condition advances with s . Thus a 24-hour forecast is produced by four applications of the 6-hour operator, not by a direct 24-hour transition.

Only the first step inherits the availability mask of the original analysis. Since each model output contains all 69 fields,

$$\mathbf{P}^{(s)} = \mathbf{1}, \quad s \geq 1. \quad (67)$$

Teacher forcing is disabled in Stage 1. In Stages 2 and 3, the next input is selected independently for each sample with probability $p_{\text{TF}} = 0.25$:

$$\mathbf{X}^{(s)} = \xi_s \mathbf{Y}^{(s)} + (1 - \xi_s) \text{stopgrad}(\widehat{\mathbf{X}}^{(s)}), \quad \xi_s \sim \text{Bernoulli}(p_{\text{TF}}), \quad (68)$$

where $\mathbf{Y}^{(s)}$ is the ground-truth state. The stop-gradient operation implements detached rollout: later-step losses update the shared transition operator at that step but do not backpropagate through earlier predicted input states. This reduces memory usage and stabilizes the transition from one-step training to self-generated multi-step inputs.

At inference time, teacher forcing is removed and every predicted state becomes the input of the next transition. The field-availability mask is set to one after the first step, and the calendar features are updated at each valid time. This inference procedure is identical for all reported lead times from three to fourteen days.

B. Data and Preprocessing Details

We used the WeatherBench2 ERA5 dataset in the 69-variable configuration. The dynamic fields are defined on a global 240×121 longitude–latitude grid and sampled at 6-hour intervals. The atmospheric component consists of five variables: geopotential height z , temperature t , specific humidity q , zonal wind u , and meridional wind v . Each of these variables is available at 13 pressure levels (50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, and 1000 hPa), yielding 65 upper-air fields in total. In addition, we include four surface variables: 2 m temperature ($t2m$), 10 m zonal wind ($u10$), 10 m meridional wind ($v10$), and mean sea-level pressure (msl), giving 69 dynamic fields overall.

We followed the data split used in our implementation: 1979–2017 for training, 2018 for validation, and 2020 for testing. Unless otherwise stated, all dynamic variables were standardized using statistics computed on the training split only. Static boundary information, including the land–sea mask and orography, was provided as additional conditioning input. Forecast lead times were expressed as multiples of 6 h, so the reported 24 h, 168 h, and 336 h results correspond to 4, 28, and 56 autoregressive steps, respectively.

C. Experimental Details

Evaluation Metrics

We adopt two standard metrics widely used in weather forecasting evaluation: the latitude-weighted root mean square error (RMSE) and the anomaly correlation coefficient (ACC). Let $\widehat{X}_{b,m,h,w}^{(s)}$ and $X_{b,m,h,w}^{(s)}$ denote the forecast and ground-truth values for sample b , field m , latitude index $h \in \{1, \dots, H\}$, and longitude index $w \in \{1, \dots, W\}$ at rollout step s , where $H = 121$ and $W = 240$ are the numbers of latitude and longitude grid cells, respectively. The latitude weight at index h with corresponding latitude ϕ_h is

$$\omega_h = \frac{\cos(\phi_h)}{H^{-1} \sum_{h'=1}^H \cos(\phi_{h'})}, \quad (69)$$

so that $\frac{1}{H} \sum_{h=1}^H \omega_h = 1$ and hence $\sum_{h=1}^H \omega_h = H$.

Root Mean Square Error. Let $e_{b,m,h,w}^{(s)} = \widehat{X}_{b,m,h,w}^{(s)} - X_{b,m,h,w}^{(s)}$ denote the per-sample forecast error of field m at step s . The latitude-weighted RMSE is

$$\text{RMSE}_m^{(s)} = \sqrt{\frac{\sum_{b,h,w} \omega_h (e_{b,m,h,w}^{(s)})^2}{\sum_{b,h,w} \omega_h}}. \quad (70)$$

Normalized RMSE is obtained by dividing the latitude-weighted RMSE of each field by the corresponding training-set standard deviation, and the reported average is computed uniformly across all 69 fields.

Anomaly Correlation Coefficient. ACC measures the anomaly correlation between forecast and observation. Let $\bar{X}_{m,h,w}$ be the climatological mean of field m at location (h, w) , and define the forecast and ground-truth anomalies as

$$\tilde{X}_{b,m,h,w}^{(s)} = \widehat{X}_{b,m,h,w}^{(s)} - \bar{X}_{m,h,w}, \quad \tilde{Y}_{b,m,h,w}^{(s)} = X_{b,m,h,w}^{(s)} - \bar{X}_{m,h,w}. \quad (71)$$

The latitude-weighted ACC is then

$$\text{ACC}_m^{(s)} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}, \quad (72)$$

where $S_{XY} = \sum_{b,h,w} \omega_h \tilde{X}_{b,m,h,w}^{(s)} \tilde{Y}_{b,m,h,w}^{(s)}$, $S_{XX} = \sum_{b,h,w} \omega_h (\tilde{X}_{b,m,h,w}^{(s)})^2$, $S_{YY} = \sum_{b,h,w} \omega_h (\tilde{Y}_{b,m,h,w}^{(s)})^2$. An ACC of 1 indicates perfect agreement, while lower values reflect reduced anomaly correlation. Throughout this paper, lower RMSE and higher ACC indicate better forecast skill.

Analysis of Learned Graph Structure and Fusion Gates

To examine whether VeinCast learns physically meaningful cross-variable interactions, we analysed the internal dynamic variable graph and the two fusion gates after training. This analysis was performed on the trained checkpoint without changing model parameters. During autoregressive inference, we registered forward hooks to collect three quantities: the dynamic adjacency matrix, the fusion feedback gate α , and the fusion query gate β .

The diagnostic was conducted on the 2020 test period using 160 samples selected to cover all four seasons. We evaluated three forecast lead times, 24, 168, and 336 hours, corresponding to short-, medium-, and long-range rollout states. For each selected sample and lead time, the learned adjacency matrix was averaged over spatial windows and then grouped by variable family. The 69 forecast fields were partitioned into geopotential height (z), temperature (t), specific humidity (q), zonal wind (u), meridional wind (v), and surface variables. Rows in the adjacency matrix denote target fields and columns denote source fields.

We further decomposed the adjacency mass according to the edge types defined by the model’s physical relation registry. These include self connections, vertical connections across pressure levels, thermodynamic–moisture links between t and q , moisture-transport links between q and u/v , wind–pressure links between z and u/v , surface–atmosphere links, and learned residual edges outside the predefined physical relation set. This decomposition tests whether the learned graph simply uses unconstrained residual links or instead allocates most of its probability mass to meteorologically interpretable relations.

The learned graph showed a structured and non-uniform interaction pattern. The strongest edge-type mass was assigned to vertical relations, with an average mass of approximately 0.46 across the analysed lead times. Learned residual edges also received a non-negligible mass of about 0.16, while self connections contributed about 0.11. The remaining physically defined cross-family relations each received smaller but consistent weights. This pattern indicates that the model relies primarily on vertical coupling across pressure levels, while retaining a residual channel to capture interactions not covered by the hand-specified physical relation types.

We also analysed the two fusion gates to understand how the latent fusion pathway contributes to different variable families. The feedback gate α controls the strength of the residual update from fusion latents back to field tokens,

$$h'_i = h_i + \alpha_i \Delta_i, \quad (73)$$

where h_i is the field token and Δ_i is the fusion-derived update. Larger α values were observed for wind and surface variables, with mean values around 0.23, whereas z and q showed lower values around 0.16. This suggests that the recurrent state update uses the fusion pathway more strongly for dynamically coupled wind and surface fields.

The query gate β controls how much the decoder query reads from the fusion representation,

$$q'_i = (1 - \beta_i) q_i + \beta_i \tilde{q}_i, \quad (74)$$

where q_i is the field-specific query and $\tilde{q}_i$ is the fusion-query representation. The query gate showed a clearer separation across variable families. The largest values appeared for wind variables, especially v and u , with mean β values of approximately 0.47 and 0.44, respectively. In contrast, geopotential

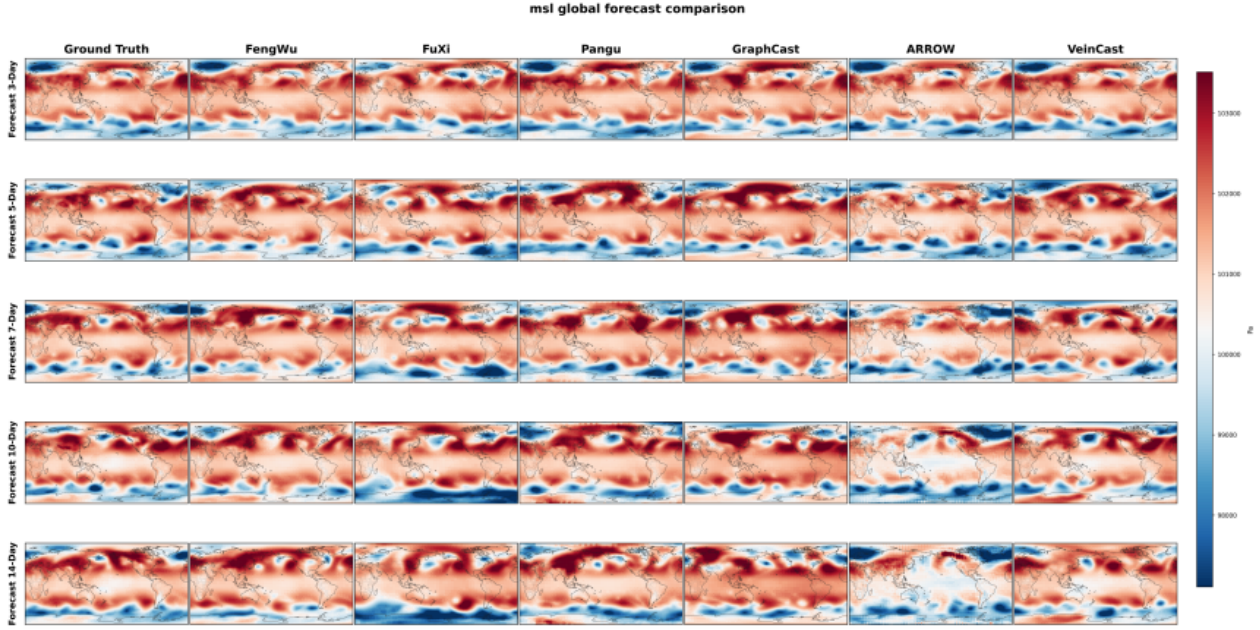


Figure 5 | Global forecast maps of VeinCast and baseline models for MSL (mean sea level pressure).

height had a much smaller mean value of approximately 0.08. These results indicate that the decoder relies more heavily on the global fusion pathway for wind-field readout, while geopotential fields are decoded mainly through field-specific representations.

Overall, this analysis shows that VeinCast does not use its graph and fusion modules as uniform black-box components. The learned adjacency emphasizes vertical atmospheric coupling, and the fusion gates allocate latent information differently across variable families. We interpret these observations as evidence that the model learns a structured internal representation aligned with major meteorological dependencies. However, this analysis is diagnostic rather than causal: the learned weights reveal how information is allocated inside the trained model, but they do not by themselves prove that each individual edge corresponds to a physical causal mechanism.

D. Results Details

RMSE and ACC of VeinCast and Baseline Models

Table 7 & 8 reports the latitude-weighted RMSE and ACC of VeinCast and the five baseline models on all remaining 300, 500, and 850 hPa upper-air variables together with MSL, complementing the selected variables presented in the main text, and the evaluation covers lead times from 3 to 14 days.

Visualization of VeinCast and Baseline Models

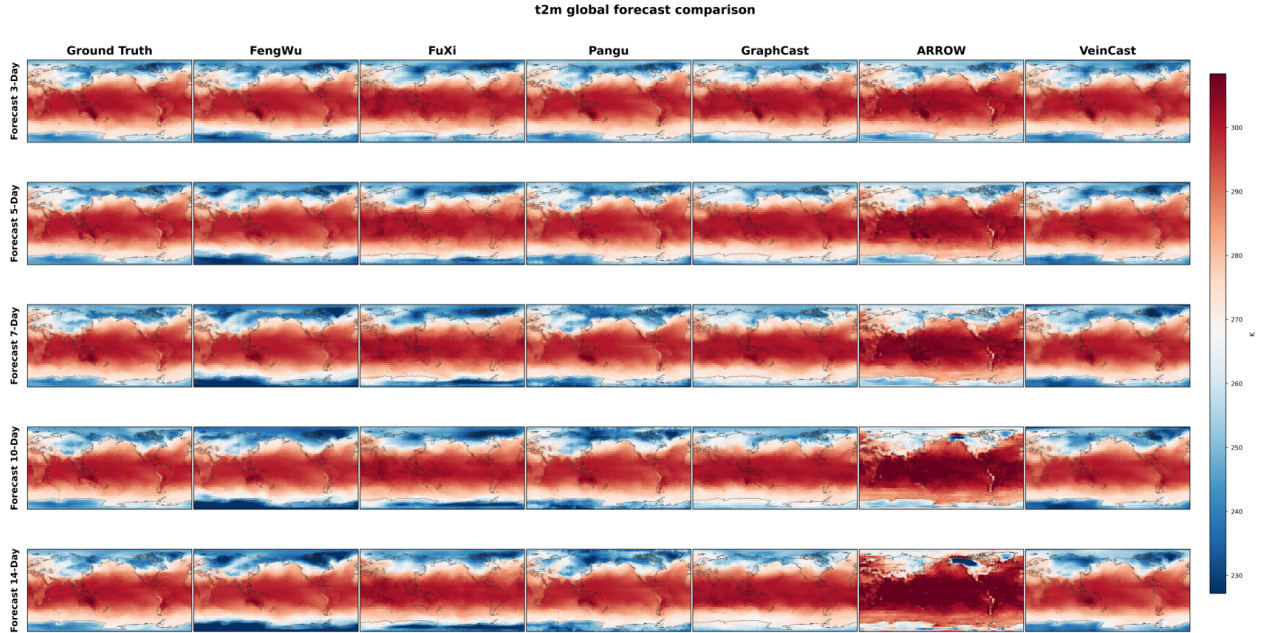
Figures 5–18 provide global forecast maps for surface variables (MSL, T2M, U10, V10) and the five upper-air variables at 500 hPa and 850 hPa (Z, T, Q, U, V). For each variable, the ground-truth ERA5 field is shown together with the forecasts of VeinCast and the five baseline models at a representative lead time. These visualizations complement the quantitative results reported in the main text and Table 7 & 8.

Table 7 | RMSE and ACC of VeinCast and baseline models on additional variables (Part 1: MSL, 300 hPa, and 500 hPa upper-air variables).

Methods		FengWu		GraphCast		Pangu-Weather		ARROW		FuXi		VeinCast	
Variable	Lead Time	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑
MSL	3-day	388.69	0.943	451.49	0.924	383.82	0.944	330.69	0.970	616.73	0.860	287.95	0.970
	5-day	605.58	0.851	722.68	0.800	556.69	0.881	580.99	0.904	877.57	0.732	504.17	0.910
	7-day	767.88	0.752	956.42	0.656	711.33	0.804	824.75	0.806	1034.71	0.641	702.64	0.828
	10-day	914.29	0.642	1189.21	0.502	860.20	0.713	1101.14	0.668	1173.20	0.552	894.67	0.724
	14-day	999.16	0.573	1345.62	0.424	950.34	0.652	1399.37	0.531	1247.11	0.521	1018.60	0.652
Z300	3-day	515.46	0.994	406.63	0.996	473.88	0.995	413.04	0.996	706.80	0.987	381.84	0.996
	5-day	885.67	0.982	832.26	0.982	813.30	0.985	745.67	0.986	1140.95	0.967	700.63	0.987
	7-day	1203.96	0.968	1235.01	0.961	1099.64	0.972	1117.16	0.967	1426.98	0.949	1009.65	0.973
	10-day	1524.87	0.951	1662.72	0.934	1382.95	0.955	1675.85	0.925	1678.41	0.931	1323.07	0.953
	14-day	1771.88	0.939	1944.76	0.914	1564.40	0.942	2394.88	0.847	1822.30	0.919	1527.72	0.937
T300	3-day	1.58	0.986	1.39	0.990	1.95	0.984	1.65	0.986	2.08	0.977	1.42	0.989
	5-day	2.25	0.973	2.27	0.974	2.76	0.968	2.41	0.970	3.01	0.954	2.00	0.979
	7-day	2.84	0.957	3.02	0.955	3.39	0.952	3.17	0.948	3.68	0.935	2.53	0.966
	10-day	3.45	0.938	3.82	0.930	3.99	0.932	4.32	0.904	4.32	0.914	3.12	0.949
	14-day	3.91	0.922	4.29	0.910	4.38	0.918	6.57	0.793	4.96	0.889	3.55	0.934
Q300	3-day	1.20e-04	0.823	1.24e-04	0.838	1.20e-04	0.812	1.12e-04	0.841	1.46e-04	0.783	1.05e-04	0.859
	5-day	1.43e-04	0.749	1.65e-04	0.721	1.42e-04	0.731	1.39e-04	0.752	1.97e-04	0.644	1.26e-04	0.799
	7-day	1.57e-04	0.691	1.97e-04	0.609	1.56e-04	0.672	1.62e-04	0.664	2.34e-04	0.528	1.42e-04	0.745
	10-day	1.70e-04	0.633	2.27e-04	0.490	1.68e-04	0.612	1.92e-04	0.538	2.64e-04	0.432	1.60e-04	0.676
	14-day	1.77e-04	0.594	2.42e-04	0.402	1.74e-04	0.579	2.38e-04	0.376	2.84e-04	0.385	1.74e-04	0.617
U300	3-day	6.53	0.929	6.08	0.939	6.89	0.921	6.43	0.935	8.84	0.870	6.08	0.938
	5-day	9.91	0.835	10.36	0.829	10.11	0.828	10.14	0.840	12.75	0.731	9.22	0.859
	7-day	12.58	0.735	13.73	0.706	12.49	0.736	13.31	0.727	15.15	0.621	11.98	0.765
	10-day	15.00	0.625	16.76	0.574	14.76	0.632	16.67	0.581	16.99	0.524	14.61	0.655
	14-day	16.57	0.548	18.53	0.487	16.08	0.566	19.56	0.450	18.28	0.459	16.33	0.577
V300	3-day	7.09	0.837	6.26	0.873	7.08	0.833	6.40	0.870	9.52	0.709	6.35	0.869
	5-day	10.81	0.621	10.88	0.629	10.57	0.618	10.26	0.670	13.62	0.404	9.80	0.688
	7-day	13.61	0.397	14.38	0.373	13.21	0.399	13.50	0.441	15.79	0.193	12.79	0.472
	10-day	15.80	0.183	17.31	0.123	15.51	0.180	16.60	0.186	17.12	0.055	15.36	0.236
	14-day	16.83	0.071	18.34	0.033	16.52	0.077	18.36	0.069	17.52	0.014	16.73	0.099
Z500	3-day	392.75	0.991	325.04	0.993	335.61	0.993	316.28	0.995	550.95	0.981	272.22	0.995
	5-day	663.65	0.976	629.48	0.975	573.01	0.980	560.84	0.982	866.36	0.955	506.64	0.984
	7-day	890.66	0.958	916.32	0.948	773.84	0.963	828.02	0.960	1069.90	0.933	733.05	0.966
	10-day	1112.65	0.936	1221.47	0.914	973.21	0.942	1212.46	0.912	1241.90	0.910	961.67	0.940
	14-day	1268.89	0.921	1430.54	0.891	1101.45	0.926	1671.97	0.839	1313.94	0.900	1109.36	0.920
T500	3-day	1.82	0.986	1.54	0.990	1.83	0.988	1.76	0.988	2.37	0.977	1.53	0.990
	5-day	2.76	0.968	2.70	0.971	2.80	0.973	2.69	0.971	3.50	0.951	2.36	0.977
	7-day	3.53	0.949	3.69	0.948	3.58	0.956	3.58	0.948	4.21	0.930	3.12	0.959
	10-day	4.27	0.928	4.63	0.920	4.33	0.937	4.69	0.907	4.84	0.910	3.84	0.938
	14-day	4.78	0.913	5.18	0.900	4.79	0.925	6.49	0.822	5.30	0.892	4.30	0.922
U500	3-day	4.65	0.923	4.49	0.930	4.76	0.919	4.49	0.931	6.35	0.861	4.30	0.935
	5-day	6.87	0.831	7.30	0.819	6.90	0.827	6.93	0.839	9.00	0.726	6.34	0.859
	7-day	8.63	0.731	9.58	0.698	8.50	0.735	9.11	0.726	10.62	0.622	8.16	0.768
	10-day	10.25	0.618	11.67	0.573	10.00	0.633	11.44	0.582	11.78	0.534	9.92	0.660
	14-day	11.26	0.541	12.82	0.506	10.89	0.569	13.46	0.465	12.45	0.488	11.04	0.583
V500	3-day	4.98	0.822	4.56	0.852	4.83	0.826	4.47	0.861	6.76	0.679	4.42	0.858
	5-day	7.38	0.605	7.48	0.609	7.10	0.609	6.99	0.663	9.40	0.381	6.61	0.680
	7-day	9.15	0.388	9.73	0.361	8.81	0.394	9.12	0.436	10.78	0.183	8.51	0.470
	10-day	10.55	0.181	11.60	0.122	10.25	0.185	11.18	0.191	11.63	0.054	10.17	0.239
	14-day	11.22	0.076	12.29	0.031	10.96	0.082	12.52	0.073	11.83	0.013	11.05	0.105

Table 8 | RMSE and ACC of VeinCast and baseline models on additional variables (Part 2: 850 hPa upper-air variables).

Methods		FengWu		GraphCast		Pangu-Weather		ARROW		FuXi		VeinCast	
Variable	Lead Time	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑	RMSE↓	ACC↑
Z850	3-day	271.29	0.975	281.23	0.974	252.22	0.979	223.05	0.986	421.01	0.940	198.25	0.987
	5-day	435.30	0.935	492.76	0.917	405.66	0.942	397.27	0.953	629.53	0.872	362.82	0.956
	7-day	565.57	0.889	684.96	0.843	532.67	0.897	569.60	0.901	764.16	0.816	516.09	0.911
	10-day	683.74	0.835	885.88	0.759	655.61	0.840	761.82	0.822	877.65	0.761	667.23	0.850
	14-day	753.93	0.799	1031.77	0.714	728.14	0.801	968.36	0.730	921.55	0.744	764.08	0.805
Q850	3-day	1.60e-03	0.927	1.74e-03	0.922	1.56e-03	0.931	1.64e-03	0.923	1.81e-03	0.905	1.43e-03	0.940
	5-day	1.91e-03	0.896	2.30e-03	0.869	1.91e-03	0.895	2.14e-03	0.871	2.26e-03	0.854	1.69e-03	0.915
	7-day	2.12e-03	0.871	2.70e-03	0.829	2.13e-03	0.869	2.58e-03	0.816	2.52e-03	0.819	1.91e-03	0.891
	10-day	2.31e-03	0.847	3.08e-03	0.789	2.32e-03	0.845	3.17e-03	0.733	2.78e-03	0.784	2.14e-03	0.864
	14-day	2.44e-03	0.831	3.28e-03	0.762	2.44e-03	0.829	3.98e-03	0.612	3.02e-03	0.747	2.31e-03	0.841
U850	3-day	3.59	0.904	3.92	0.893	3.57	0.905	3.45	0.917	5.02	0.821	3.35	0.918
	5-day	5.07	0.806	5.93	0.760	5.01	0.807	5.22	0.812	6.85	0.680	4.75	0.836
	7-day	6.16	0.711	7.46	0.632	6.04	0.715	6.74	0.689	7.90	0.586	5.93	0.746
	10-day	7.05	0.614	8.70	0.515	6.91	0.623	8.29	0.539	8.69	0.507	6.99	0.647
	14-day	7.54	0.555	9.39	0.455	7.39	0.569	9.80	0.406	9.13	0.471	7.64	0.578
V850	3-day	3.69	0.792	3.79	0.786	3.53	0.802	3.32	0.840	4.98	0.626	3.34	0.828
	5-day	5.20	0.577	5.67	0.533	4.96	0.589	5.07	0.634	6.55	0.346	4.75	0.649
	7-day	6.27	0.377	7.04	0.303	5.96	0.396	6.47	0.415	7.32	0.176	5.91	0.455
	10-day	7.05	0.203	8.09	0.109	6.78	0.221	7.72	0.205	7.82	0.070	6.87	0.262
	14-day	7.42	0.123	8.48	0.036	7.17	0.137	8.67	0.096	7.93	0.036	7.38	0.151


Figure 6 | Global forecast maps of VeinCast and baseline models for T2M (2 m temperature).

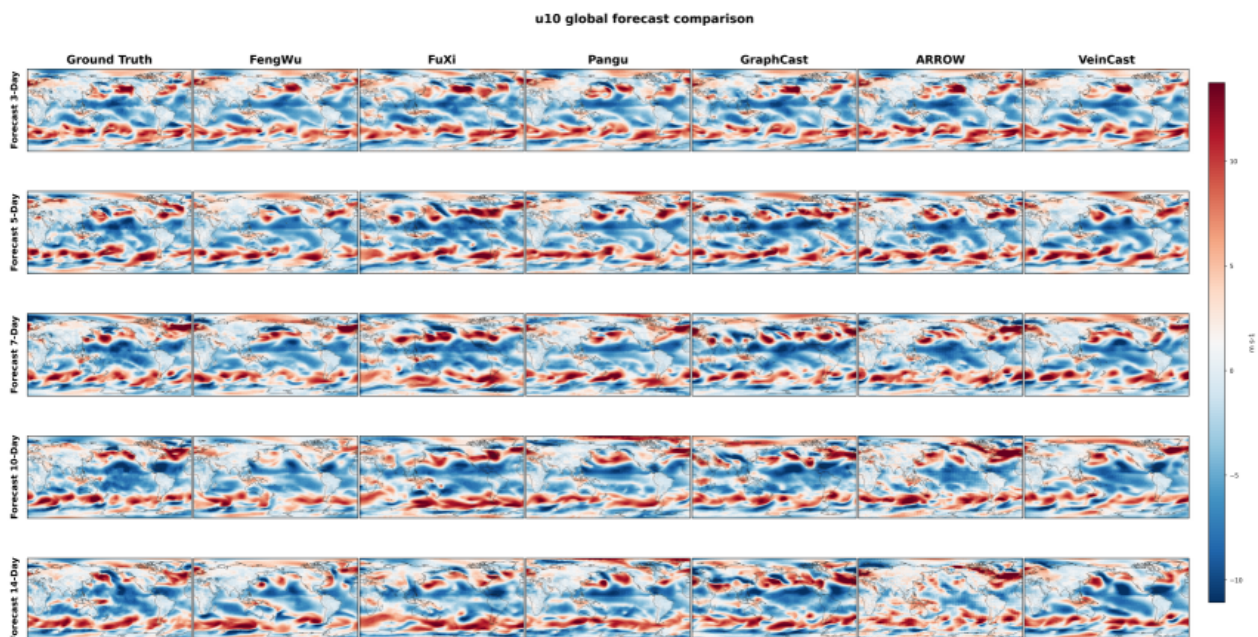


Figure 7 | Global forecast maps of VeinCast and baseline models for U10 (10 m zonal wind).

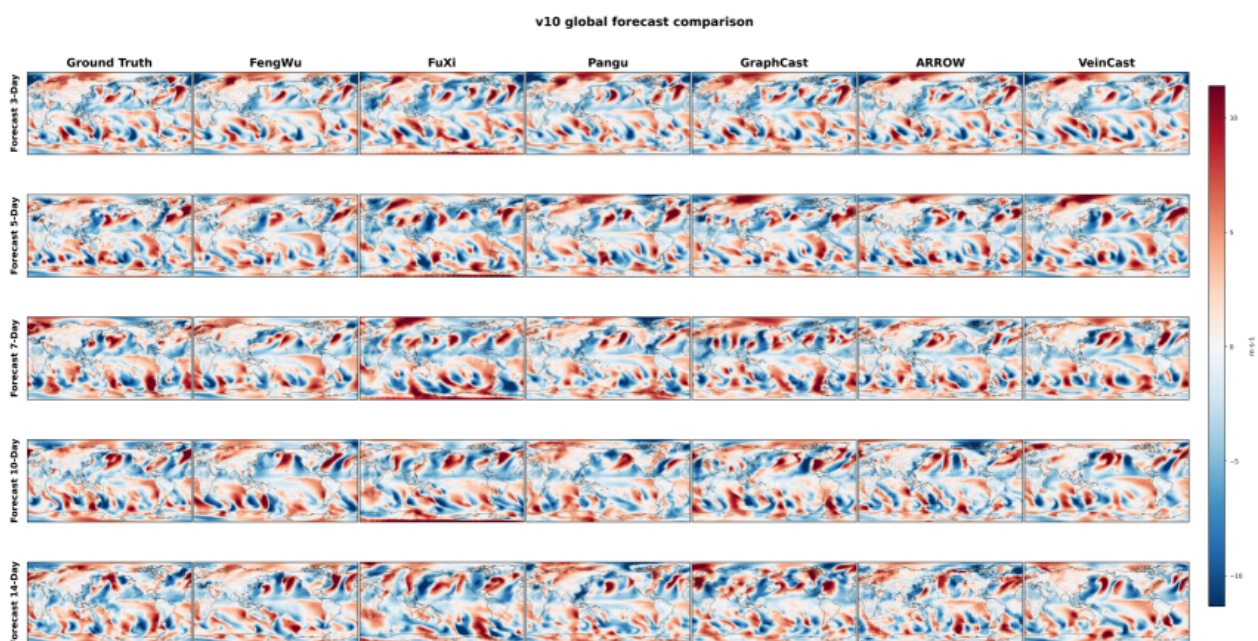


Figure 8 | Global forecast maps of VeinCast and baseline models for V10 (10 m meridional wind).

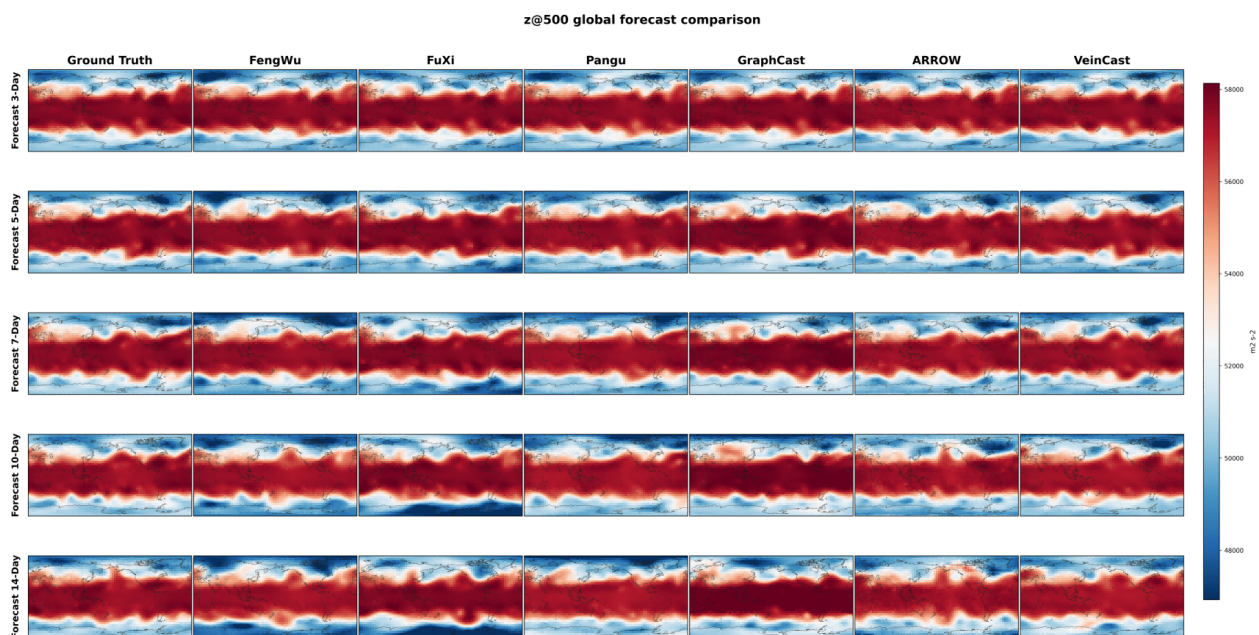


Figure 9 | Global forecast maps of VeinCast and baseline models for Z500 (geopotential at 500 hPa).

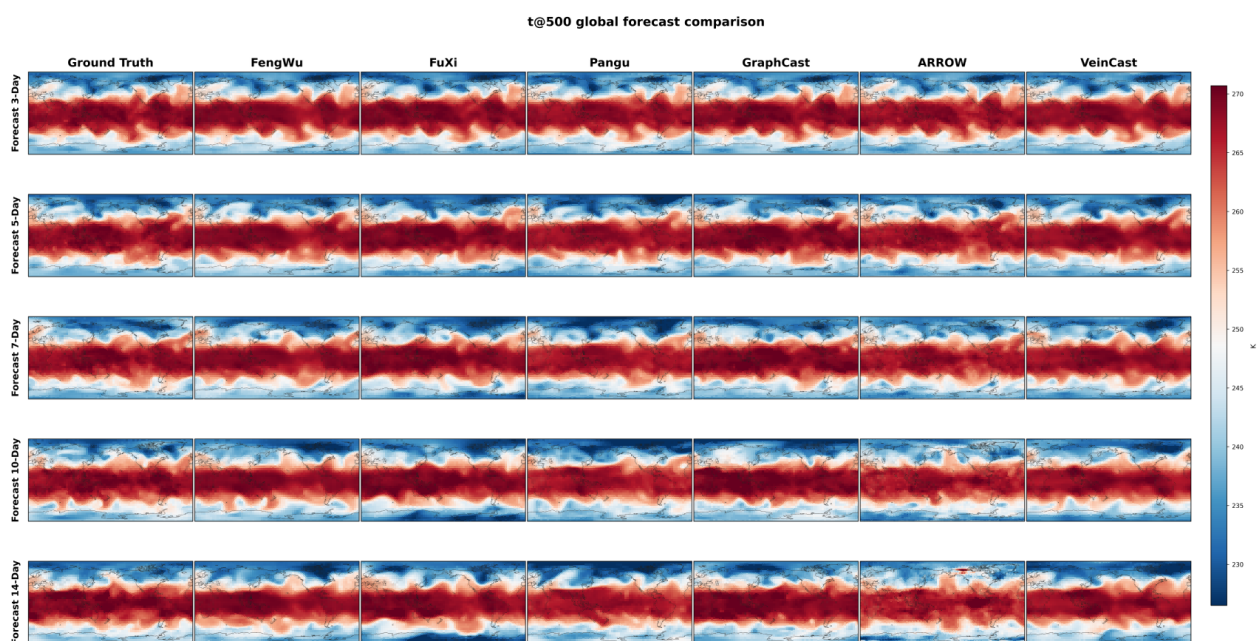


Figure 10 | Global forecast maps of VeinCast and baseline models for T500 (temperature at 500 hPa).

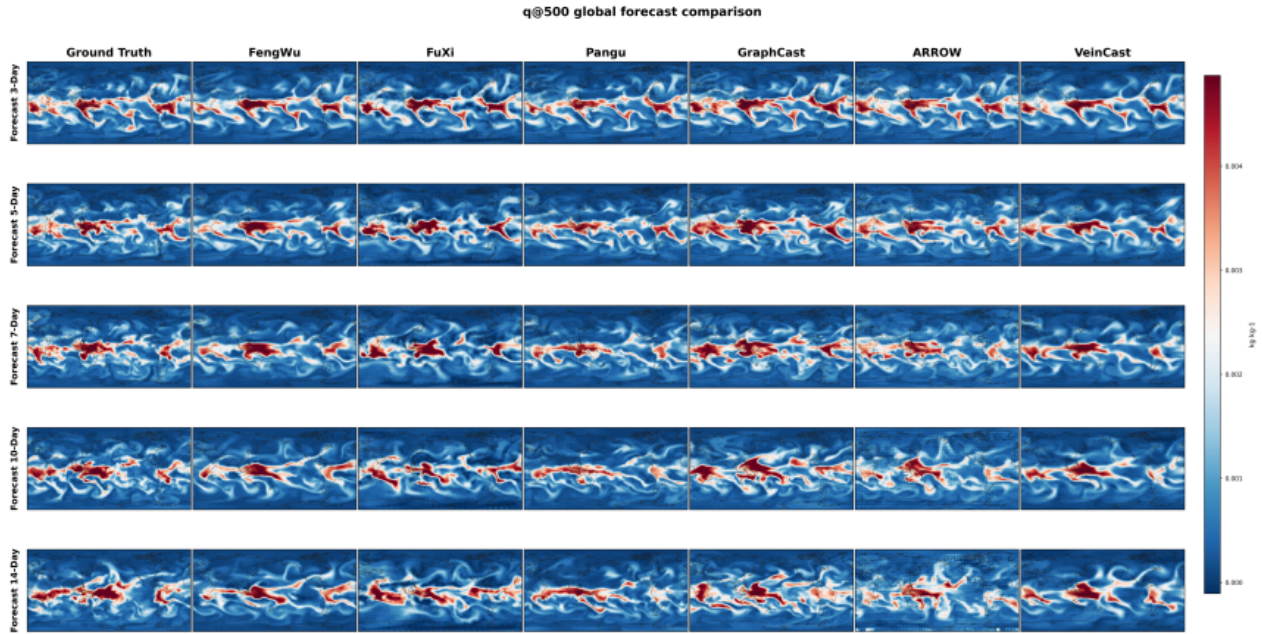


Figure 11 | Global forecast maps of VeinCast and baseline models for Q500 (specific humidity at 500 hPa).

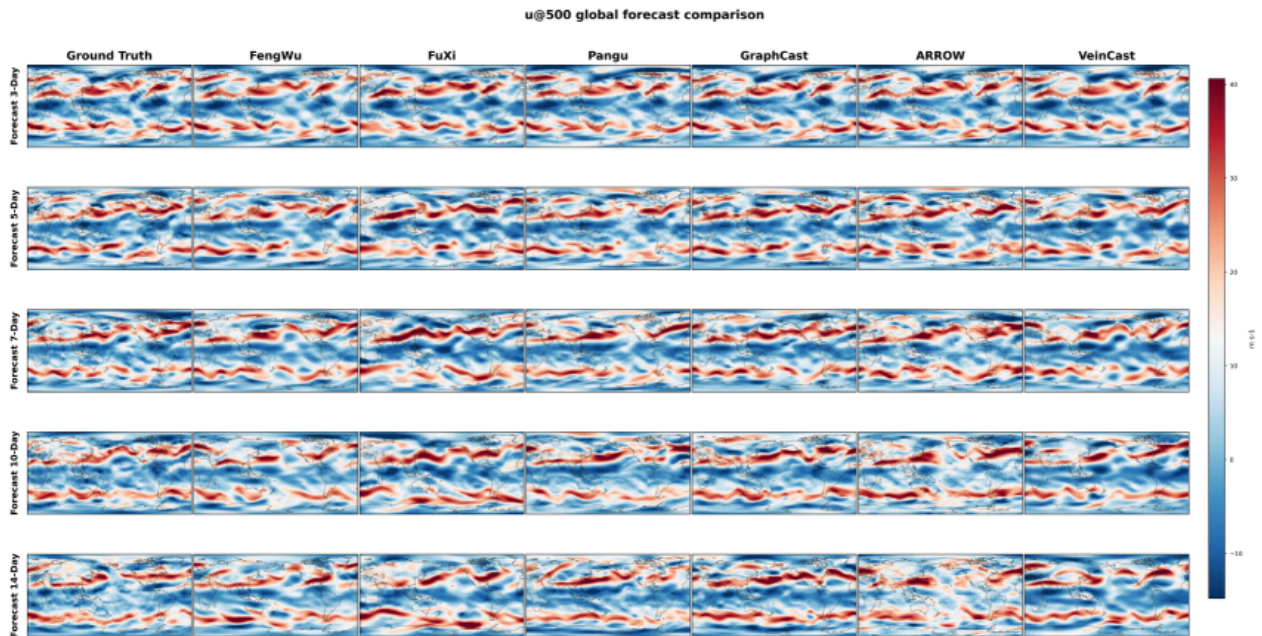


Figure 12 | Global forecast maps of VeinCast and baseline models for U500 (zonal wind at 500 hPa).

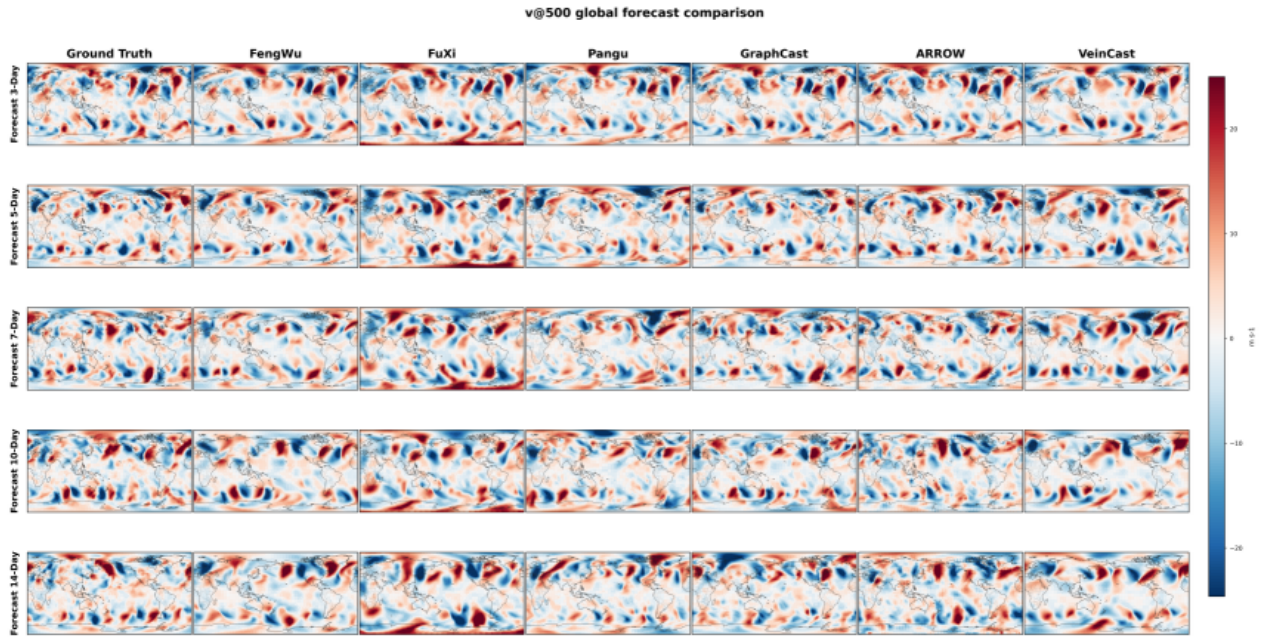


Figure 13 | Global forecast maps of VeinCast and baseline models for V500 (meridional wind at 500 hPa).

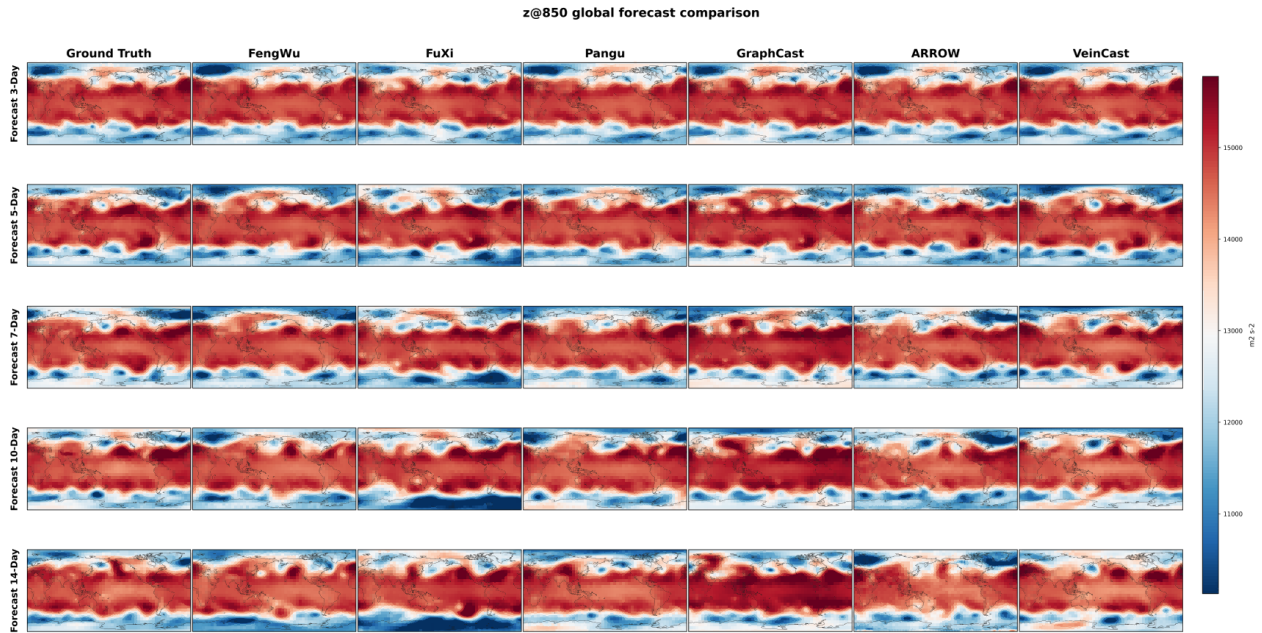


Figure 14 | Global forecast maps of VeinCast and baseline models for Z850 (geopotential at 850 hPa).

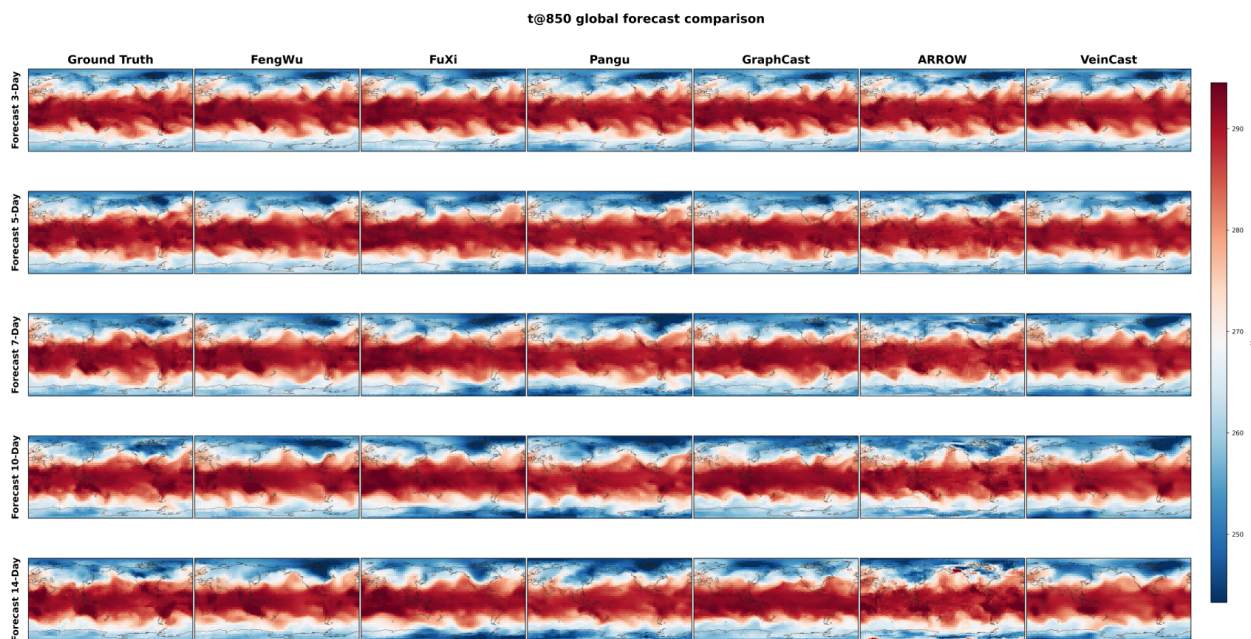


Figure 15 | Global forecast maps of VeinCast and baseline models for T850 (temperature at 850 hPa).

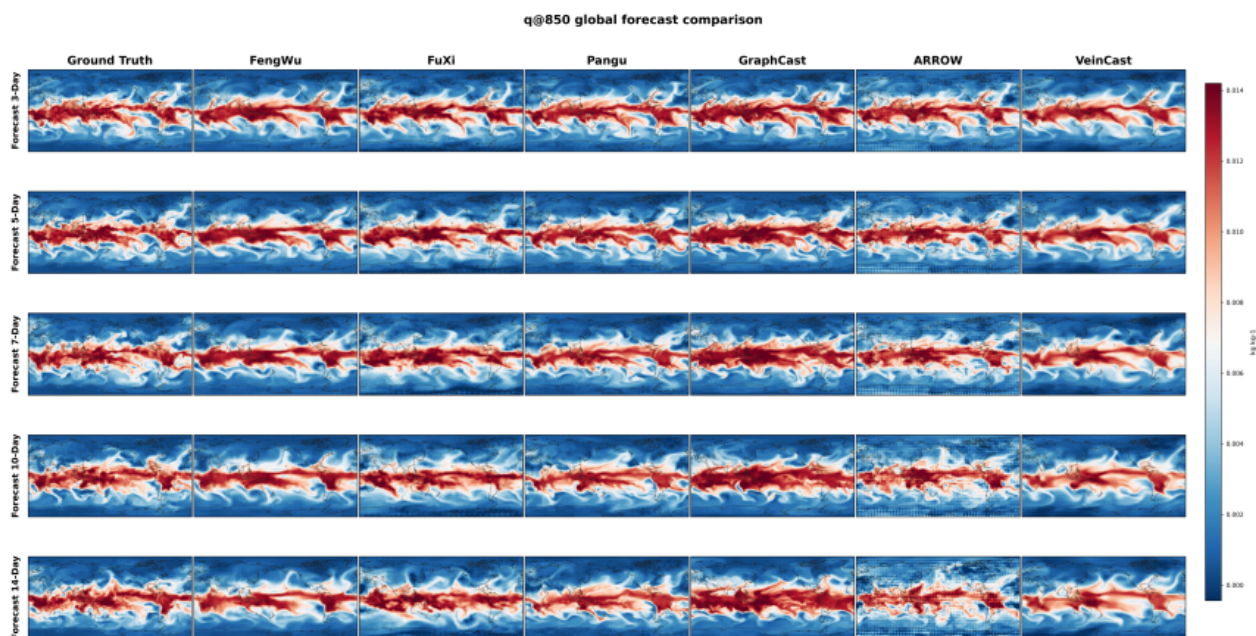


Figure 16 | Global forecast maps of VeinCast and baseline models for Q850 (specific humidity at 850 hPa).

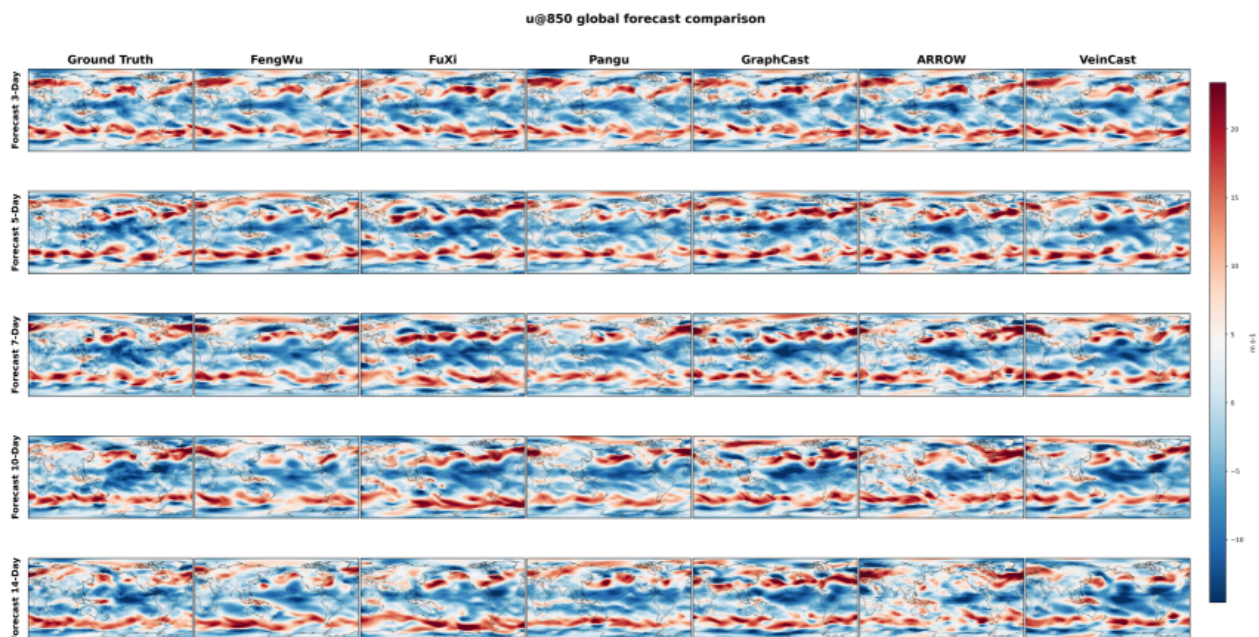


Figure 17 | Global forecast maps of VeinCast and baseline models for U850 (zonal wind at 850 hPa).

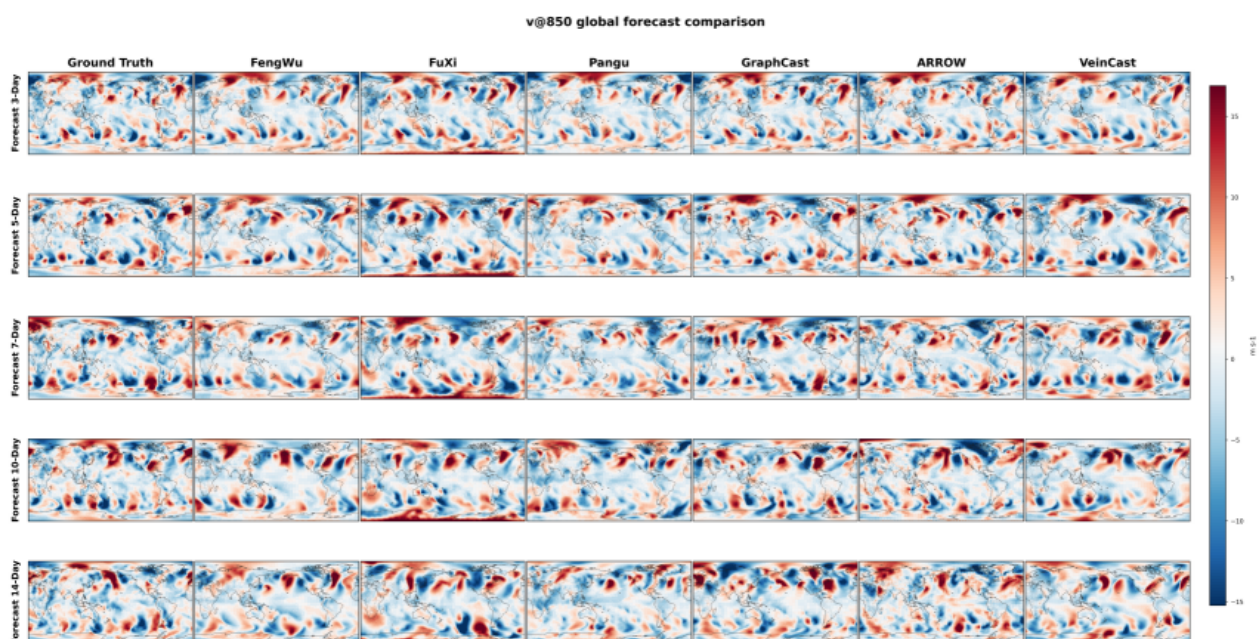


Figure 18 | Global forecast maps of VeinCast and baseline models for V850 (meridional wind at 850 hPa).