WINDWEAVER: ALONG THE WIND AND ACROSS SCALES FOR GENERATIVE WEATHER SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Fine-resolution weather sequences support applications sensitive to local conditions and their evolution. Generative super-resolution must recover unresolved fine-scale states and connect their changes across gaps in coarse observations. Coarse wind offers approximate cross-time correspondence while leaving multivariate fine-scale evolution undetermined. We introduce WindWeaver, which uses atmospheric transport to structure the computation of generative updates. Coarse wind organizes cross-time feature exchange, while the network learns how the exchanged information contributes to multivariate reconstruction. Wind-derived coordinates retrieve features of the current candidate sequence at each flow evaluation. Separate readouts predict coarse block means and within-block detail, while a locally correlated source initializes unresolved variation. These interfaces support velocity-based conditional flow matching (CFM) and a task-adapted, denoiser-based adaptive flow matching (AFM) host. Across East Asian ERA5, same-product CERRA, and paired ERA5-to-CERRA reconstruction, WindWeaver improves probabilistic reconstruction of wind, temperature, cloud, and radiation in both hosts. Controlled CFM experiments establish net communication gains over branch removal in state and six-hour increment scores. Matched static and reversed correspondence show that the coordinate rule matters. Increment benefits extend across every reconstructed variable group. These results support using approximate atmospheric correspondence to guide generative feature exchange for reconstructing multivariate weather states and their temporal changes.

1 INTRODUCTION

Fine-resolution weather information supports applications such as local wind and solar resource assessment (Stengel et al., 2020; Springenberg et al., 2026). Temporal resolution matters alongside spatial resolution: coarse output intervals can leave rapid weather changes unresolved (Leinonen et al., 2024; Ingstad et al., 2026). Spatial downscaling recovers structure within coarse grid cells, while temporal refinement fills gaps between available snapshots. Joint spatiotemporal super-resolution therefore requires fine-scale fields and the changes connecting them. Multiple fine-scale sequences can match the same coarse observations, motivating conditional generation to represent this unresolved variability (Leinonen et al., 2021; Rozet & Louppe, 2023; Schmidt et al., 2025). The objective is to reconstruct a distribution of weather sequences, including both their spatial detail and their intervening changes.

Generative modeling has expanded the tools available for this reconstruction. Adversarial and diffusion approaches recover fine-scale atmospheric variability (Stengel et al., 2020; Mardani et al., 2025). Flow matching learns vector fields that transform a source distribution into a conditional target distribution (Lipman et al., 2023; Tong et al., 2024). Weather downscaling with flow matching and adaptive flow matching (AFM) demonstrates this framework's promise for small-scale reconstruction (Delefosse et al., 2026; Fotiadis et al., 2025). Recurrent and video generative models capture temporal dependencies (Leinonen et al., 2021; Srivastava et al., 2024), while trajectory priors support multivariate reconstruction through posterior sampling (Schmidt et al., 2025; Aich et al., 2026). Building on these sequence-modeling capabilities, we ask how available physical relations can organize generative updates to unresolved weather states.

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

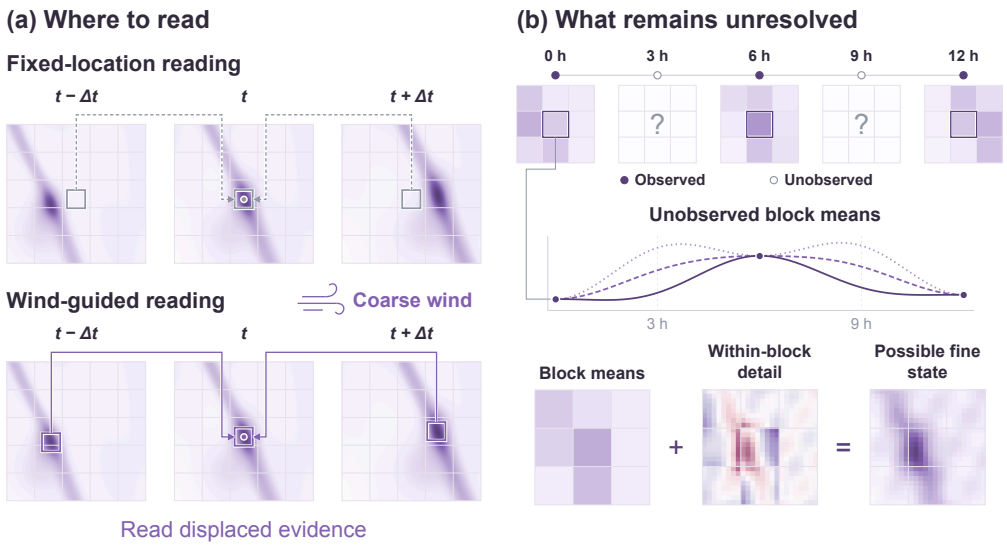

Figure 1: **Physical correspondence for unresolved weather states.** (a) Coarse wind supplies candidate locations for reading related features at neighboring weather times. (b) Observed aggregates leave intermediate block means and within-block detail undetermined. The fields and curves are schematic; instantaneous observations at 0/6/12 h leave states at 3/9 h to be reconstructed.

Atmospheric transport provides a useful but incomplete relation for this purpose. A moving structure can leave related information at different locations across weather times. Following motion to locate upstream information is central to semi-Lagrangian atmospheric computation and precipitation nowcasting (Staniforth & Côté, 1991; Pulkkinen et al., 2019). In super-resolution, coarse wind supplies approximate correspondence between fine-scale states that are themselves unknown. Transport-based predictors combine motion with learned intensity changes or source variations (Zhang et al., 2023; Verma et al., 2024). Physical knowledge also guides latent dynamics (Le Guen & Thome, 2020) and reconstruction constraints (Harder et al., 2023; Debeire et al., 2026). These uses motivate a specific question for conditional generation: *can an incomplete physical relation organize useful generative information exchange when its physical variables are already available as inputs?*

We introduce **WindWeaver**, which uses atmospheric transport to structure the computation of generative updates. Coarse wind organizes cross-time feature exchange, while the network learns how the exchanged information contributes to multivariate reconstruction. Motion-guided exchange within generation has precedents in video models such as FLATTEN and Motion-I2V (Cong et al., 2024; Shi et al., 2024). At each flow evaluation, wind-derived locations retrieve features from neighboring times in the current candidate sequence, and a learned mixer combines the retrieved content. Atmospheric transport guides feature access; the network learns the resulting multivariate update without prescribing an advection law for the outputs.

WindWeaver couples this communication operator with interfaces for the uncertainty left by coarse observations. Winds at three pressure levels route latent features into separate readouts for coarse block means and zero-mean within-block detail. Both remain learnable at missing times, where even the intermediate coarse states are unresolved (Figure 1). A locally correlated source combines short temporal memory with independent random inputs at each weather time. Communication, readouts, and source organize information exchange, prediction scales, and initial variability at distinct stages of reconstruction. These interfaces integrate with a velocity-based conditional flow matching (CFM) host and a task-adapted denoiser-based AFM host.

Experiments evaluate this design on East Asian ERA5, same-product CERRA, and paired ERA5-to-CERRA reconstruction. WindWeaver improves probabilistic reconstruction of wind, temperature, cloud, and radiation in both hosts. Controlled CFM comparisons retain the same physical inputs and backbone attention. Adding wind-guided exchange improves balanced state CRPS and six-hour increment scores over branch removal. Static and reversed correspondence produce higher balanced

state and increment CRPS than branch removal. Net increment CRPS and RMSE gains extend across all four variable groups. These comparisons link reconstruction gains to the tested communication design and correspondence rule under matched physical inputs.

Our contributions are: **(i)** a pressure-level wind-guided communication operator that gives generative weather updates a transport-informed computational structure; **(ii)** complementary source and observation-scale prediction interfaces that support this operator in two flow parameterizations; and **(iii)** controlled evidence for net multivariate reconstruction gains and the role of correspondence when reconstructing temporal changes. Together, these findings identify coarse-wind correspondence as a useful guide for feature exchange in the evaluated multivariate weather reconstruction tasks.

2 Related Work

Generative spatiotemporal weather reconstruction. Probabilistic downscaling represents unresolved atmospheric variability through adversarial generation (Stengel et al., 2020), residual diffusion (Mardani et al., 2025), and flow matching (Delefosse et al., 2026). AFM learns a conditional source and adaptive noise scale (Fotiadis et al., 2025). Wind applications include paired ERA5-to-CERRA reconstruction (Merizzi et al., 2024) and downscaling across resolutions and forecast lead times (Springenberg et al., 2026). Recurrent stochastic models and video diffusion address temporal dependencies (Leinonen et al., 2021; Srivastava et al., 2024). Score-based data assimilation infers trajectories from incomplete observations (Rozet & Louppe, 2023); weather models use learned sequence priors for multivariate reconstruction (Schmidt et al., 2025; Aich et al., 2026). ModAFNO and HourGlass interpolate between weather states (Leinonen et al., 2024; Ingstad et al., 2026), while FlowConvGRU combines temporal interpolation with spatial refinement (Harder et al., 2023). These methods establish several routes to temporal reconstruction. WindWeaver examines the additional value of organizing generative feature exchange through available atmospheric motion.

Physical information and reconstruction constraints. Physical knowledge informs learning objectives, representations, and architecture (Karniadakis et al., 2021). Dynamical-generative downscaling combines regional climate simulation with diffusion-based refinement (Lopez-Gomez et al., 2025). Transport-based neural models couple learned motion to physical evolution (de Bézenac et al., 2018; Verma et al., 2024). PhyDNet and physics-integrated variational autoencoders combine physical components with learned complementary factors (Le Guen & Thome, 2020; Takeishi & Kalousis, 2021). ClimPhyDM combines learned transport and source terms to model weather evolution through dynamics matching (Singh et al., 2026). For atmospheric super-resolution, GeoFAR combines geography-informed and frequency-aware representations (Xu et al., 2026a), while PISR uses multiscale objectives based on hydrostatic primitive equations (Xu et al., 2026b). Hard-constrained downscaling preserves coarse aggregates (Harder et al., 2023); PC-AFM adds soft conservation constraints while managing their interaction with the generative objective (Debeire et al., 2026). WindWeaver uses approximate transport correspondence to route latent features within each generative update. Learned mixing determines their contribution to multivariate reconstruction, alongside observation-scale constraints.

Motion-guided communication within generation. Motion correspondence connects features across locations and times. TrajGRU learns location-dependent recurrent connections (Shi et al., 2017); BasicVSR and BasicVSR++ use optical flow for video-feature alignment and propagation (Chan et al., 2021; 2022). Atmospheric nowcasting combines motion with stochastic extrapolation in pysteps (Pulkkinen et al., 2019) and physical evolution with conditional generation in NowcastNet (Zhang et al., 2023). Within diffusion, FLATTEN organizes attention along source-video optical-flow trajectories (Cong et al., 2024). Motion-I2V predicts motion from a reference image and text, then incorporates warped reference features into temporal attention (Shi et al., 2024). WindWeaver shares this use of motion to structure generative computation. Its correspondence comes from coarse winds at multiple pressure levels and retrieves features of the current, unresolved candidate weather sequence. Our branch-removal, static, and reversed-correspondence controls test the net benefit of this exchange and its coordinate rule under matched physical inputs and backbone attention.

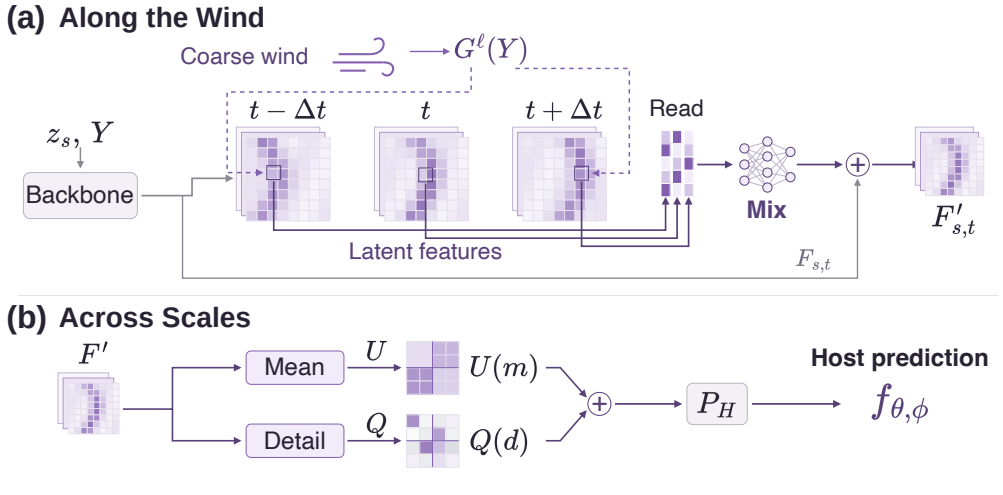

Figure 2: **One WindWeaver predictor evaluation.** (a) **Along the Wind:** coarse-wind coordinates $G^\ell(Y)$ locate neighboring-time candidate features; the mixer combines these reads with current-time features in a residual update. (b) **Across Scales:** scale readouts combine $U(m)$ and zero-block-mean $Q(d)$ before task projection P_H. Figure 4 shows host integration.

Source structure and prediction scales. The source distribution and predictor parameterization also shape the reconstruction problem. Correlated stochasticity appears in atmospheric parameterization (Berner et al., 2017), video diffusion (Ge et al., 2023), and spatial flow-matching sources (Zarei & Kalayeh, 2026). WindWeaver combines local correlation and short temporal memory with independent per-time random inputs. For prediction across scales, LapFlow generates Laplacian pyramid components (Zhao et al., 2026), while DDNM refines nullspace content under fixed observations (Wang et al., 2023). Our readouts separate coarse block means from within-block detail according to the observation grid. Both components remain variable at missing times, and this spatial decomposition differs from CorrDiff's conditional-mean/residual decomposition (Mardani et al., 2025). These interfaces support wind-guided communication while retaining flexibility in the unresolved sequence.

3 WindWeaver

WindWeaver retrieves features from neighboring weather times at locations determined by coarse wind, then combines them through a learned residual mixer. Scale-separated readouts convert the updated features into predictions; a locally correlated source initializes the sequence (Figure 2).

3.1 Task and generative state

We reconstruct nine channels in $X \in \mathbb{R}^{C\times T\times H\times W}$: zonal and meridional wind at 850, 700, and 300 hPa, 2 m temperature (T2m), total cloud cover, and surface solar radiation. Conditions Y contain five six-hourly coarse observations plus geographic and calendar features. Targets span 0–12 h at hourly resolution for East Asian ERA5 and three-hourly resolution for CERRA. Weather time t indexes states within this sequence; generation time $s \in [0, 1]$ indexes refinement of the entire candidate.

We generate normalized residuals around a deterministic coarse-input anchor $a(Y)$:

$$z_\star = \mathcal{P}_H S^{-1}(X - a(Y)), \qquad \widehat{X} = a(Y) + Sz_1. \tag{1}$$

The diagonal matrix S contains training-estimated channel scales. In controlled coarsening, $\mathcal{P}_H$ projects residuals into the observation nullspace, preserving the aggregates supplied by the anchor. For paired ERA5→CERRA, $\mathcal{P}_H = I$ permits target-product mean corrections. Appendix B.1 defines the instantaneous and radiation-interval constraints.

3.2 WIND-GUIDED LATENT COMMUNICATION

Following semi-Lagrangian backward sampling (Staniforth & Côté, 1991), coarse wind supplies approximate correspondence between neighboring weather times. Spatial and temporal self-attention encode the candidate and conditions into $F_{s,t}$; a learned projection produces three feature groups, one per pressure level. For source time $\tau = t \pm \Delta t$, where Δt is the target time step, the sampling coordinates are

$$G^{\ell}_{t\leftarrow\tau}(p) = p - (t-\tau)J(p)^{-1}\overline{\boldsymbol{u}}^{\ell}_{t,\tau}(p), \qquad \overline{\boldsymbol{u}}^{\ell}_{t,\tau} = \tfrac{1}{2}(\boldsymbol{u}^{\ell}_{t} + \boldsymbol{u}^{\ell}_{\tau}). \tag{2}$$

Here ℓ indexes pressure levels, winds are interpolated from coarse observations, and $t-\tau$ is in seconds. The metric $J(p)$ converts feature-grid displacement to east/north meters, accounting for grid geometry and encoder stride. Bilinear reads from the three source-time feature groups at $G^{\ell}_{t\leftarrow\tau}$ are concatenated into $\widetilde{F}_{s,\tau\to t}$. A pointwise mixer combines these with current-time projected features $F^{\mathrm{proj}}_{s,t}$:

$$F'_{s,t} = F_{s,t} + \mathcal{M}_{\phi}\big([\widetilde{F}_{s,t-\Delta t\to t}, F^{\mathrm{proj}}_{s,t}, \widetilde{F}_{s,t+\Delta t\to t}]\big). \tag{3}$$

Coordinates depend only on Y and can be cached; features are recomputed from z_s at every predictor call. Wind guides feature access, and the network learns the output update. Appendix B.3 specifies mixer initialization and boundary handling.

3.3 COMPLEMENTARY PREDICTION ACROSS SCALES

Separate readouts predict missing-time coarse states and within-block structure. Let D average blocks with area weights and U replicate block values onto the fine grid, so $DU = I$. With $\mathcal{Q} = I - UD$, the prediction is

$$f_{\theta,\phi}(z_s, s; Y) = \mathcal{P}_H[U(m_{\theta,\phi}(F')) + \mathcal{Q}(d_{\theta,\phi}(F'))], \qquad D\mathcal{Q} = 0. \tag{4}$$

Before observation projection, block averaging recovers m and detail has zero block mean. The split can represent any field. Both readouts depend on the random candidate, allowing missing-time block means to vary across members. The separate projection $\mathcal{P}_H$ constrains supplied aggregates.

3.4 LOCALLY CORRELATED SOURCE

The source combines local persistence with independent random inputs at each weather time. For white-noise tensors ϵ, η drawn independently, spatial filter K, and short temporal filter T_ρ, define

$$\xi = \mathcal{E}_{\mathcal{P}_H\epsilon}\Big(\mathcal{P}_H\Big[\tfrac{1}{2}\epsilon + \tfrac{\sqrt{3}}{2}T_\rho K\eta\Big]\Big). \tag{5}$$

The operation $\mathcal{E}_b$ matches each sample/channel's whole-window squared norm to that of b. Filtering introduces fixed-grid correlation; ϵ supplies fresh inputs at each weather time before projection and normalization. Each ensemble member evolves from one complete source draw. Appendix B.4 specifies the filters and normalization.

3.5 SHARED INTERFACES FOR TWO FLOW HOSTS

CFM initializes $z_0 = \xi$; task-adapted AFM uses $z_0 = \bar{\mu}_\psi(Y) + \sigma_z\xi$, retaining its learned conditional center $\bar{\mu}_\psi = \mathcal{P}_H\mu_\psi$ and adaptive scalar scale σ_z (Fotiadis et al., 2025). Both follow the straight training path $z_s = (1-s)z_0 + sz_\star$, with velocities

$$\widehat{v}(z_s, s; Y) = \begin{cases} f_{\theta,\phi}(z_s, s; Y), & \text{CFM}, \\ [\mathcal{D}_{\theta,\phi}(z_s, s; Y) - z_s]/(1-s), & \text{AFM}. \end{cases} \tag{6}$$

Equation (4) predicts CFM velocity directly or supplies the trainable branch of AFM's EDM-preconditioned denoiser $\mathcal{D}$ (Karras et al., 2022).

Each host jointly trains its predictor, mixer, and readouts by velocity regression, weighting wind, temperature, cloud, and radiation equally in normalized residual coordinates. AFM additionally regularizes its center and calibrates σ_z on training windows. Sampling applies N projected Euler updates, $z^{k+1} = \mathcal{P}_H[z^k + \widehat{v}(z^k, k/N; Y)/N]$, then Eq. (1) recovers physical fields. Hosts are trained separately on the same backbone family; full objectives and algorithms are in Appendix B.

4 EXPERIMENTS

4.1 PROTOCOLS AND EVALUATION

We use three protocols: East Asian ERA5 at $1.5°$/6 h to $0.25°$/1 h; same-product CERRA at 22 km/6 h to 5.5 km/3 h; and paired ERA5–CERRA with ERA5 inputs on 33 km conditioning centers at 6 h and CERRA targets at 5.5 km/3 h (Hersbach et al., 2020; Ridal et al., 2024; Copernicus Climate Change Service, 2022a;b). The first two recover artificially aggregated fields; the third also accommodates product differences. All reconstruct the nine channels in Section 3; spatial contexts are specified in Appendix B. We evaluate both CFM and AFM with and without the complete WindWeaver design, denoted Full and Base, respectively. Within each host, comparisons share inputs, anchor, projection, attention backbone, training budget, and checkpoint selection.

Training uses 2020–2021. In each month of 2022, validation starts on day 2 at 00 UTC and testing on day 16 at 00/12 UTC (72 validation and 144 test windows). We use three training seeds, 12 ensemble members, and 30 flow evaluations (NFE). Checkpoint selection minimizes validation balanced continuous ranked probability score (CRPS): four variable-group scores, normalized by training residual scales and equally weighted. Wind, T2m, and cloud are scored at missing times; radiation uses daylight intervals. Root mean square error (RMSE) scores the ensemble mean; Appendix B.8 defines metrics and masks (Gneiting & Raftery, 2007). Seed SD measures training variation; paired month-block intervals condition on the three fitted models. Appendix E reports validation curves and runtime costs.

4.2 MULTIVARIATE GAINS IN TWO FLOW HOSTS

Table 1: **Multivariate gains in both hosts.** Positive percentages denote CRPS reductions from matched Base. Balanced CRPS: three-seed mean ± sample SD; lower is better.

		Variable CRPS reduction (%)				Balanced CRPS		Reduction
Protocol	Host	Wind	T2m	Cloud	Rad.	Base	WindWeaver	(%)
East Asian ERA5	CFM	11.23	9.63	3.76	3.32	0.3042 ± 0.0002	0.2830 ± 0.0002	6.96
	AFM	15.04	11.63	6.37	4.60	0.3292 ± 0.0003	0.2977 ± 0.0005	9.55
Same-product CERRA	CFM	16.22	5.81	9.69	7.32	0.4899 ± 0.0009	0.4347 ± 0.0008	11.27
	AFM	19.61	11.23	11.46	10.05	0.5301 ± 0.0021	0.4530 ± 0.0003	14.54
ERA5 → CERRA	CFM	18.81	5.04	9.52	9.86	0.4888 ± 0.0013	0.4309 ± 0.0003	11.85
	AFM	5.49	9.73	6.27	5.55	0.5013 ± 0.0005	0.4679 ± 0.0015	6.67

Reductions use detailed variable scores: normalized groups for same-product CERRA, physical units otherwise. Balanced normalization is protocol-specific.

WindWeaver improves all four variable groups over Base in every protocol–host pair, reducing balanced CRPS by 6.67–14.54% (Table 1); ensemble-mean RMSE improves in all six comparisons. Appendix C reports full state and increment results; Appendix D maps spatial gains and remaining errors.

4.3 THE VALUE OF CORRESPONDENCE WITHIN COMMUNICATION

CFM controls separate the branch's net benefit from the choice of correspondence. No-exchange removes the added branch, preserving the source, readouts, and backbone attention. Independently trained Static and Reverse retain Full's architecture and parameter count, with zero and negated displacements, respectively. Their scores include the resulting boundary effects. Balanced-CRPS means follow Full $<$ No-exchange $<$ Static $<$ Reverse in all three protocols (Table 2).

The matched MLP improves East Asian ERA5 balanced CRPS by 0.19%, compared with WindWeaver's 6.96%. The gain from this capacity control remains close to Base.

Table 2: **Matched correspondence and interface controls.** CFM balanced CRPS (three-seed mean ± SD). Appendix C reports AFM and source controls.

CFM configuration	East Asian ERA5	Same-product CERRA	ERA5 → CERRA
Base	0.3042 ± 0.0002	0.4899 ± 0.0009	0.4888 ± 0.0013
WindWeaver	0.2830 ± 0.0002	0.4347 ± 0.0008	0.4309 ± 0.0003
No-exchange	0.2864 ± 0.0003	0.4612 ± 0.0115	0.4570 ± 0.0006
Static correspondence	0.2979 ± 0.0008	0.4715 ± 0.0009	0.4654 ± 0.0011
Reverse correspondence	0.2983 ± 0.0006	0.4773 ± 0.0011	0.4682 ± 0.0006
Capacity control	0.3036 ± 0.0003	0.4885 ± 0.0023	0.4897 ± 0.0005
Scale readouts only	0.2993 ± 0.0001	0.4860 ± 0.0034	0.4818 ± 0.0013
w/o scale readouts	0.2885 ± 0.0003	0.4400 ± 0.0010	0.4494 ± 0.0011
w/o structured source	0.2944 ± 0.0001	0.4471 ± 0.0009	0.4653 ± 0.0024

Table 3: **Net gains in six-hour change reconstruction.** CFM increment CRPS in physical units and Full reductions relative to No-exchange, Static and Reverse; net gains include paired 95% intervals.

		Increment CRPS ↓		Full reduction (%)		
Protocol	Variable	Full	No-exchange	Net gain [95% interval]	vs Static	vs Reverse
East Asian ERA5	Wind	0.754	0.768	1.77 [1.34, 2.18]	2.07	7.87
	T2m	0.569	0.570	0.17 [0.10, 0.24]	0.21	5.13
	Cloud	0.0973	0.0979	0.63 [0.40, 0.90]	0.73	2.77
	Radiation	28.627	28.661	0.12 [0.07, 0.17]	0.16	2.46
Same-product CERRA	Wind	1.245	1.431	12.98 [10.78, 14.51]	14.57	15.57
	T2m	0.528	0.542	2.43 [1.86, 2.99]	3.39	4.21
	Cloud	0.1145	0.1190	3.79 [2.88, 4.46]	4.56	5.88
	Radiation	15.960	16.527	3.43 [2.74, 4.32]	4.07	5.29
ERA5 → CERRA	Wind	1.387	1.607	13.71 [12.08, 15.90]	19.24	21.41
	T2m	0.652	0.665	1.96 [1.57, 2.38]	2.34	5.00
	Cloud	0.1485	0.1536	3.37 [2.71, 4.16]	4.06	4.95
	Radiation	39.185	39.930	1.87 [1.50, 2.36]	2.17	6.86

Both instantaneous endpoints are missing; radiation uses valid daylight interval means. Units: m/s, K, fraction, W/m^2. Appendix C gives endpoint pairs, seed SDs, RMSE, Base comparisons and interval details.

4.4 Net gains in changes between unobserved states

We score signed increments $\Delta_\tau X^{(m)} = X^{(m)}(t+\tau) - X^{(m)}(t)$ against reference increments, preserving member pairing across weather times. Both endpoints of each six-hour instantaneous-variable pair are missing; radiation compares daylight interval-mean fluxes (Table 3).

In CFM, wind-guided communication improves six-hour increment CRPS over No-exchange for every variable group in all three protocols, with all paired 95% intervals above zero (Table 3). Increment RMSE also improves throughout; Tables 12–14 report all 24 CRPS/RMSE comparisons.

Increment-CRPS means follow the same ordering across all variable groups and protocols. Figure 3 visualizes increment errors for No-exchange, Static, Reverse, and Full in one wind case. Gains can reflect both state accuracy and temporal dependence.

4.5 Supporting source and prediction scales

Removing scale readouts or the structured source worsens balanced CRPS in every protocol for both hosts (Table 2; Appendix C). Full also outperforms scale readouts alone. These component effects are measured within the full design; the appendix includes source-memory and spatial-smoothing controls.

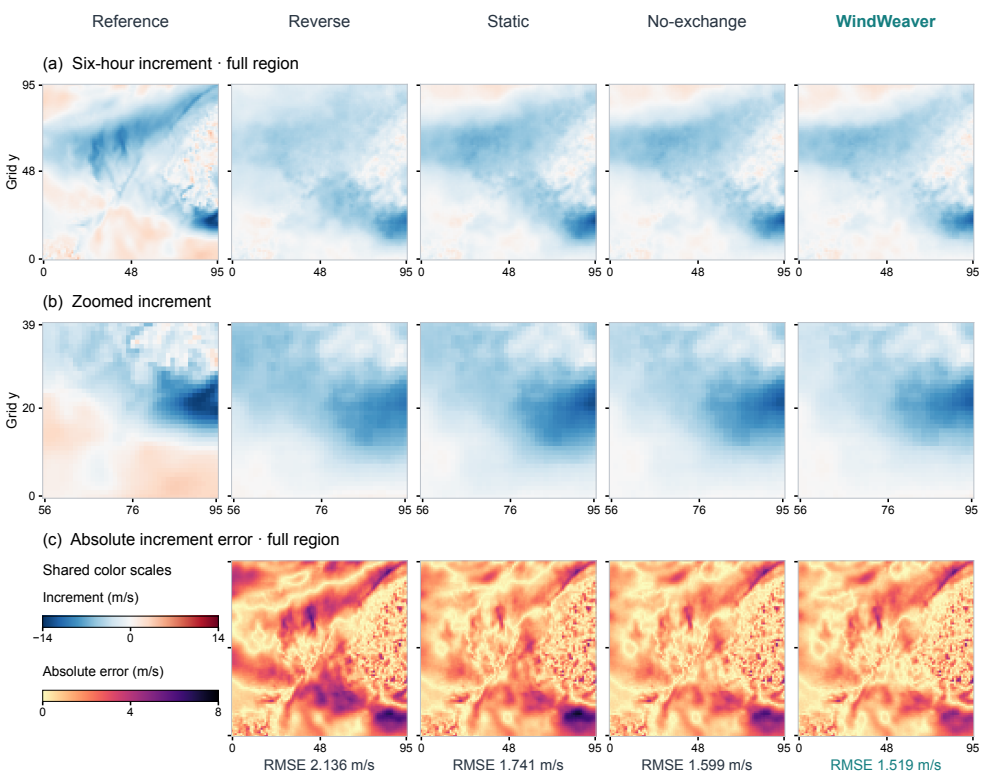

Figure 3: **Reconstructing changes between missing states.** Same-product CERRA u_{850} on 16 January 2022: 21 UTC minus 15 UTC, using seed-42 CFM ensemble means. (a) Six-hour increments. (b) A shared 40×40-cell crop selected by maximum reference-increment variance. (c) Absolute increment errors and weighted RMSE over the full 96×96 region. WindWeaver has lower case RMSE than all three controls. Increment and error maps use separate shared m/s scales; fine reference structure remains unresolved.

Relative to Base, the full design improves missing-time block-mean CRPS, detail CRPS, and wind spatial-increment CRPS for both hosts on East Asian ERA5 and ERA5→CERRA (Table 22). Figures 5–7 show spatially varying gains and remaining errors.

5 Discussion

WindWeaver shows that coarse wind can contribute to reconstruction through the way it organizes computation, even when the same wind fields are already available as inputs. Coarse observations provide approximate read locations, and each flow evaluation recomputes feature content from the current candidate sequence. The communication geometry remains tied to observed motion while the exchanged representations evolve during generation. This lets jointly reconstructed states provide context for one another throughout refinement. Motion-guided exchange has precedents in video generation (Cong et al., 2024; Shi et al., 2024); our experiments establish its utility for coarse-wind-guided communication among unresolved multivariate weather states in the evaluated tasks.

The controlled CFM comparisons separate the net value of communication from the value of its retrieval geometry. Full improves balanced state CRPS and six-hour increment CRPS over No-exchange, which retains the source, readouts, wind inputs, and backbone attention. Static and Reverse retain the communication architecture but change its read coordinates; both score worse than No-exchange in these comparisons (Tables 2 and 3). The added branch therefore benefits reconstruction with wind-guided retrieval, and the tested alternatives show that its value depends on the correspondence rule, including boundary handling. For instantaneous variables, both endpoints of

the six-hour comparisons are unobserved. These gains consequently extend to changes between jointly inferred states at unobserved times.

Benefits for temperature, cloud, and radiation show how a shared correspondence can serve variables with different evolution mechanisms. CFM Full improves their increment CRPS and RMSE over No-exchange in all three protocols (Tables 12–14). The wind relation acts on shared latent features, leaving variable-specific changes to the learned reconstruction. These gains support using transport to locate useful context across variables, even though transport describes only part of their evolution. The network learns how that context contributes to each output. This gives a limited physical relation a concrete role in multivariate reconstruction.

Communication supplies context while leaving substantial reconstruction freedom. At missing times, both block means and within-block details remain unknown; scale-separated readouts represent both components explicitly. The source supplies initial variability through local correlations and independent random inputs at each weather time. These interfaces assign distinct computational roles to feature exchange, prediction scales, and stochastic initialization. Component ablations and coarse/detail diagnostics support their contributions within the complete design, with effects varying by protocol (Table 2; Appendix C.6). Integration into velocity-CFM and denoiser-AFM demonstrates reuse across two prediction parameterizations.

The evidence covers retrospective reconstruction with two flow parameterizations and one backbone family on the evaluated monthly test dates. Both hosts support complete-design gains and state-level component comparisons; net temporal and correspondence-rule evidence comes from CFM. Increment scores reflect both state accuracy and temporal dependence, leaving direct trajectory recovery and full-sequence calibration for further evaluation. A natural next question is how the predictor should weigh retrieved context when coarse-wind correspondence becomes more uncertain. Testing this relationship across broader weather coverage and observation resolutions would clarify when transport-guided communication remains useful and how its use should adapt to the available physical information.

6 Conclusion

Weather super-resolution requires reconstructing fine-scale states together with the changes between sparsely observed times. We introduced WindWeaver to make coarse atmospheric transport useful during this joint reconstruction. At each flow evaluation, wind-derived correspondence connects features of neighboring states in the current candidate sequence, and a learned mixer uses this context to compute multivariate updates. Scale-separated readouts and a locally correlated source support this communication within velocity-CFM and denoiser-AFM.

Across East Asian ERA5, same-product CERRA, and paired ERA5–CERRA reconstruction, the complete design improves probabilistic reconstruction of all four variable groups in both hosts. Controlled CFM experiments clarify the value of communication when the same wind inputs and backbone attention are already available. Wind-guided exchange yields net improvements in missing-state reconstruction and six-hour changes; static and reversed correspondence show that the organization of this exchange matters. Increment gains extend to temperature, cloud, and radiation, supporting the use of wind-derived correspondence in a shared multivariate representation.

Together, these findings give a concrete answer to the question motivating this work: even when atmospheric motion is observed only at coarse resolution, it can guide useful communication among weather states that must be inferred together. Coarse wind specifies where related context is sought, while the network learns how that context contributes to variables with different evolution mechanisms. In the evaluated tasks, this division makes atmospheric transport useful for reconstruction without prescribing the complete evolution of each output. WindWeaver provides a practical example of embedding physical knowledge in the computation of generative updates. We hope this work encourages further investigation of how observable physical relationships can help generative models reconstruct unresolved states and their changes.

AI USE STATEMENT

Generative AI tools assisted with literature search and synthesis, hypothesis refinement, and discussions of experimental design. They also supported code implementation and debugging, data processing, interpretation of results, figure preparation, manuscript structuring, and polishing. Numerical results are drawn from executed experiments and their recorded outputs. The authors are responsible for verifying AI-assisted material and for the final text, code, data provenance, references, and scientific claims.

REPRODUCIBILITY STATEMENT

Section 3 specifies the model and host interfaces. The appendices detail observation operators, source construction, losses, training and sampling, data alignment, validation-based model selection, evaluation masks, and runtime costs. Result tables summarize three training seeds; paired month-block intervals quantify uncertainty in temporal-score improvements conditional on the fitted models. Model comparisons use matched inputs and training budgets within each host.

Anonymous code repository: `https://anonymous.4open.science/r/WindWeaver/`

ETHICS STATEMENT

This study uses atmospheric reanalysis products and involves no human participants or personal data. Reconstructed fields are model-generated estimates that include unresolved variability. Operational or safety-critical use requires further validation beyond the retrospective reconstruction experiments reported here.

REFERENCES

Michael Aich, Andreas Fürst, Florian Sestak, Carlos Ruiz-Gonzalez, Niklas Boers, and Johannes Brandstetter. WIND: Weather inverse diffusion for zero-shot atmospheric modeling. *arXiv preprint arXiv:2602.03924*, 2026. URL https://arxiv.org/abs/2602.03924.

Judith Berner, Ulrich Achatz, Lauriane Batté, Lisa Bengtsson, Alvaro de la Cámara, Hannah M. Christensen, Matteo Colangeli, Danielle R. B. Coleman, et al. Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, 98:565–588, 2017. doi: 10.1175/BAMS-D-15-00268.1. URL https://doi.org/10.1175/BAMS-D-15-00268.1.

Kelvin C. K. Chan, Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. BasicVSR: The search for essential components in video super-resolution and beyond. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4947–4956, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Chan_BasicVSR_The_Search_for_Essential_Components_in_Video_Super-Resolution_and_CVPR_2021_paper.html.

Kelvin C. K. Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. BasicVSR++: Improving video super-resolution with enhanced propagation and alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5972–5981, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Chan_BasicVSR_Improving_Video_Super-Resolution_With_Enhanced_Propagation_and_Alignment_CVPR_2022_paper.html.

Yuren Cong, Mengmeng Xu, Christian Simon, Shoufa Chen, Jiawei Ren, Yanping Xie, Juan-Manuel Perez-Rua, Bodo Rosenhahn, Tao Xiang, and Sen He. FLATTEN: Optical flow-guided attention for consistent text-to-video editing. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2310.05922.

Copernicus Climate Change Service. CERRA sub-daily regional reanalysis data for Europe on pressure levels from 1984 to present. Climate Data Store, dataset, 2022a. URL https://cds.climate.copernicus.eu/datasets/reanalysis-cerra-pressure-levels.

Copernicus Climate Change Service. CERRA sub-daily regional reanalysis data for Europe on single levels from 1984 to present. Climate Data Store, dataset, 2022b. URL https://cds.climate.copernicus.eu/datasets/reanalysis-cerra-single-levels.

Emmanuel de Bézenac, Arthur Pajot, and Patrick Gallinari. Deep learning for physical processes: Incorporating prior scientific knowledge. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=By4HsfWAZ.

Kevin Debeire, Aytaç Paçal, Pierre Gentine, Luis Medrano-Navarro, Nils Thuerey, and Veronika Eyring. Physics-constrained adaptive flow matching for climate downscaling. *arXiv preprint arXiv:2604.03459*, 2026. URL https://arxiv.org/abs/2604.03459.

Aymeric Delefosse, Anastase Charantonis, and Dominique Béréziat. Super-resolving coarse-resolution weather forecasts with flow matching. *Environmental Data Science*, 5:e25, 2026. doi: 10.1017/eds.2026.10046. URL https://www.cambridge.org/core/journals/environmental-data-science/article/superresolving-coarseresolution-weather-forecasts-with-flow-matching/2DB59047E39B2F650407CAEEDE1686A7.

Christopher A. T. Ferro. Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1917–1923, 2014. doi: 10.1002/qj.2270. URL https://doi.org/10.1002/qj.2270.

Stathi Fotiadis, Noah D. Brenowitz, Tomas Geffner, Yair Cohen, Michael Pritchard, Arash Vahdat, and Morteza Mardani. Adaptive flow matching for resolving small-scale physics. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 17489–17521. PMLR, 2025. URL https://proceedings.mlr.press/v267/fotiadis25a.html.

Songwei Ge, Seungjun Nah, Guilin Liu, Tyler Poon, Andrew Tao, Bryan Catanzaro, David Jacobs, Jia-Bin Huang, Ming-Yu Liu, and Yogesh Balaji. Preserve your own correlation: A noise prior for video diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 22930–22941, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Ge_Preserve_Your_Own_Correlation_A_Noise_Prior_for_Video_Diffusion_ICCV_2023_paper.html.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437. URL https://doi.org/10.1198/016214506000001437.

Paula Harder, Alex Hernandez-Garcia, Venkatesh Ramesh, Qidong Yang, Prasanna Sattegeri, Daniela Szwarcman, Campbell Watson, and David Rolnick. Hard-constrained deep learning for climate downscaling. *Journal of Machine Learning Research*, 24(365):1–40, 2023. URL https://jmlr.org/papers/v24/23-0158.html.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146 (730):1999–2049, 2020. doi: 10.1002/qj.3803. URL https://doi.org/10.1002/qj.3803.

Magnus Sikora Ingstad, Mariana C. A. Clare, Olav Ersland, Vera Gahlen, Håvard Homleid Haugen, Ophélia Miralles, Even M. Nordhagen, Thomas N. Nipen, Ivar A. Seierstad, John Bjørnar Bremnes, Michael Maier-Gerber, Zied Ben Bouallègue, Harrison Cook, Christian Lessig, Gert Mertes, Cathal O'Brien, Florian Pinault, Ana Prieto Nemesio, and Matthew Chantry. HourGlass: A probabilistic data-driven temporal downscaler for global and regional weather forecasting. *arXiv preprint arXiv:2607.11457*, 2026. URL https://arxiv.org/abs/2607.11457.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021. doi: 10.1038/s42254-021-00314-5. URL https://doi.org/10.1038/s42254-021-00314-5.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/a98846e9d9cc01cfb87eb694d946ce6b-Abstract.html.

Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11474–11484, 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Le_Guen_Disentangling_Physical_Dynamics_From_Unknown_Factors_for_Unsupervised_Video_Prediction_CVPR_2020_paper.html.

Jussi Leinonen, Daniele Nerini, and Alexis Berne. Stochastic Super-Resolution for Downscaling Time-Evolving Atmospheric Fields With a Generative Adversarial Network. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7211–7223, 2021. doi: 10.1109/TGRS.2020.3032790. URL https://arxiv.org/abs/2005.10374.

Jussi Leinonen, Boris Bonev, Thorsten Kurth, and Yair Cohen. Modulated adaptive fourier neural operators for temporal interpolation of weather forecasts. *arXiv preprint arXiv:2410.18904*, 2024. URL https://arxiv.org/abs/2410.18904.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2210.02747.

Ignacio Lopez-Gomez, Zhong Yi Wan, Leonardo Zepeda-Núñez, Tapio Schneider, John Anderson, and Fei Sha. Dynamical-generative downscaling of climate model ensembles. *Proceedings of the National Academy of Sciences*, 122(17):e2420288122, 2025. doi: 10.1073/pnas.2420288122. URL https://doi.org/10.1073/pnas.2420288122.

Morteza Mardani, Noah Brenowitz, Yair Cohen, Jaideep Pathak, Chieh-Yu Chen, Cheng-Chin Liu, Arash Vahdat, Mohammad Amin Nabian, Tao Ge, Akshay Subramaniam, Karthik Kashinath, Jan Kautz, and Mike Pritchard. Residual corrective diffusion modeling for km-scale atmospheric downscaling. *Communications Earth & Environment*, 6:124, 2025. doi: 10.1038/s43247-025-02042-5. URL https://doi.org/10.1038/s43247-025-02042-5.

Fabio Merizzi, Andrea Asperti, and Stefano Colamonaco. Wind speed super-resolution and validation: From era5 to cerra via diffusion models. *arXiv preprint arXiv:2401.15469*, 2024. URL https://arxiv.org/abs/2401.15469.

Seppo Pulkkinen, Daniele Nerini, Andrés A. Pérez Hortal, Carlos Velasco-Forero, Alan Seed, Urs Germann, and Loris Foresti. Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geoscientific Model Development*, 12:4185–4219, 2019. doi: 10.5194/gmd-12-4185-2019. URL https://gmd.copernicus.org/articles/12/4185/2019/.

Martin Ridal, Eric Bazile, Patrick Le Moigne, Roger Randriamampianina, Semjon Schimanke, Ulf Andrae, Lars Berggren, Pierre Brousseau, Per Dahlgren, Lisette Edvinsson, et al. CERRA, the copernicus european regional reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, 150(763):3385–3411, 2024. doi: 10.1002/qj.4764. URL https://doi.org/10.1002/qj.4764.

François Rozet and Gilles Louppe. Score-based data assimilation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 40521–40541, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/7f7fa581cc8a1970a4332920cdf87395-Abstract-Conference.html.

Michael Scheuerer and Thomas M. Hamill. Variogram-Based Proper Scoring Rules for Probabilistic Forecasts of Multivariate Quantities. *Monthly Weather Review*, 143(4):1321–1334, 2015. doi: 10.1175/MWR-D-14-00269.1. URL https://doi.org/10.1175/MWR-D-14-00269.1.

Jonathan Schmidt, Luca Schmidt, Felix M. Strnad, Nicole Ludwig, and Philipp Hennig. A generative framework for probabilistic, spatiotemporally coherent downscaling of climate simulation. *npj Climate and Atmospheric Science*, 8:270, 2025. doi: 10.1038/s41612-025-01157-y. URL https://www.nature.com/articles/s41612-025-01157-y.

Xiaoyu Shi, Zhaoyang Huang, Fu-Yun Wang, Weikang Bian, Dasong Li, Yi Zhang, Manyuan Zhang, Ka Chun Cheung, Simon See, Hongwei Qin, Jifeng Dai, and Hongsheng Li. Motion-I2V: Consistent and controllable image-to-video generation with explicit motion modeling. In *ACM SIGGRAPH 2024 Conference Papers*, 2024. URL https://arxiv.org/abs/2401.15977.

Xingjian Shi, Zhihan Gao, Leonard Lausen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Deep learning for precipitation nowcasting: A benchmark and a new model. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1706.03458.

Gurjeet Sangra Singh, Frantzeska Lavda, and Alexandros Kalousis. Climate physics dynamic matching. *arXiv preprint arXiv:2608.26907*, 2026. URL https://arxiv.org/abs/2608.26907.

Maximilian Springenberg, Noelia Otero, Yuxin Xue, and Jackie Ma. DiffScale: Continuous downscaling and bias correction of subseasonal wind speed forecasts using diffusion models. *Journal of Advances in Modeling Earth Systems*, 18(3):e2025MS005282, 2026. doi: 10.1029/2025MS005282. URL https://doi.org/10.1029/2025MS005282.

Prakhar Srivastava, Ruihan Yang, Gavin Kerrigan, Gideon Dresdner, Jeremy McGibbon, Christopher Bretherton, and Stephan Mandt. Precipitation Downscaling with Spatiotemporal Video Diffusion. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://openreview.net/forum?id=hhnkH8ex5d.

Andrew Staniforth and Jean Côté. Semi-lagrangian integration schemes for atmospheric models—a review. *Monthly Weather Review*, 119(9):2206–2223, 1991. URL https://journals.ametsoc.org/view/journals/mwre/119/9/1520-0493_1991_119_2206_slisfa_2_0_co_2.xml.

Karen Stengel, Andrew Glaws, Dylan Hettinger, and Ryan N. King. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117: 16805–16815, 2020. doi: 10.1073/pnas.1918964117. URL https://doi.org/10.1073/pnas.1918964117.

Naoya Takeishi and Alexandros Kalousis. Physics-integrated variational autoencoders for robust and interpretable generative modeling. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/7ca57a9f85a19a6e4b9a248c1daca185-Abstract.html.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. URL https://arxiv.org/abs/2302.00482.

Yogesh Verma, Markus Heinonen, and Vikas Garg. ClimODE: Climate and weather forecasting with physics-informed neural ODEs. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/21f3c05fc5afcaebd8a71d7e98a1b171-Abstract-Conference.html.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=mRieQgMtNTQ.

Chang Xu, Gencer Sumbul, Li Mi, Robin Zbinden, and Devis Tuia. GeoFAR: Geography-informed frequency-aware super-resolution for climate data. In *International Conference on Learning Representations*, 2026a. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/039c30e9af8039fbd1b58da9d04f38e9-Abstract-Conference.html.

Chang Xu, Gencer Sumbul, Hugo Porta, Manon Béchaz, Sebastian Schemm, and Devis Tuia. Physics-informed super-resolution of atmospheric data. *arXiv preprint arXiv:2607.18877*, 2026b. URL https://arxiv.org/abs/2607.18877.

Arman Zarei and Mahdi M. Kalayeh. Spatially-grounded flow matching: Structured source distributions for image generation. *arXiv preprint arXiv:2608.15452*, 2026. URL https://arxiv.org/abs/2608.15452.

Yuchen Zhang, Mingsheng Long, Kaiyuan Chen, Lanxiang Xing, Ronghua Jin, Michael I. Jordan, and Jianmin Wang. Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, 619: 526–532, 2023. doi: 10.1038/s41586-023-06184-4. URL https://www.nature.com/articles/s41586-023-06184-4.

Zelin Zhao, Petr Molodyk, Haotian Xue, and Yongxin Chen. LapFlow: Laplacian multi-scale flow matching for generative modeling. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/c919a2b5ec1de69f2629f9119676e336-Abstract-Conference.html.

A Integration with the Flow Hosts

Figure 4 shows how the predictor in Figure 2 enters each host. Both hosts share the source, geometry and readout interfaces and jointly train their predictor and host parameters. Appendix B gives the equations and algorithms.

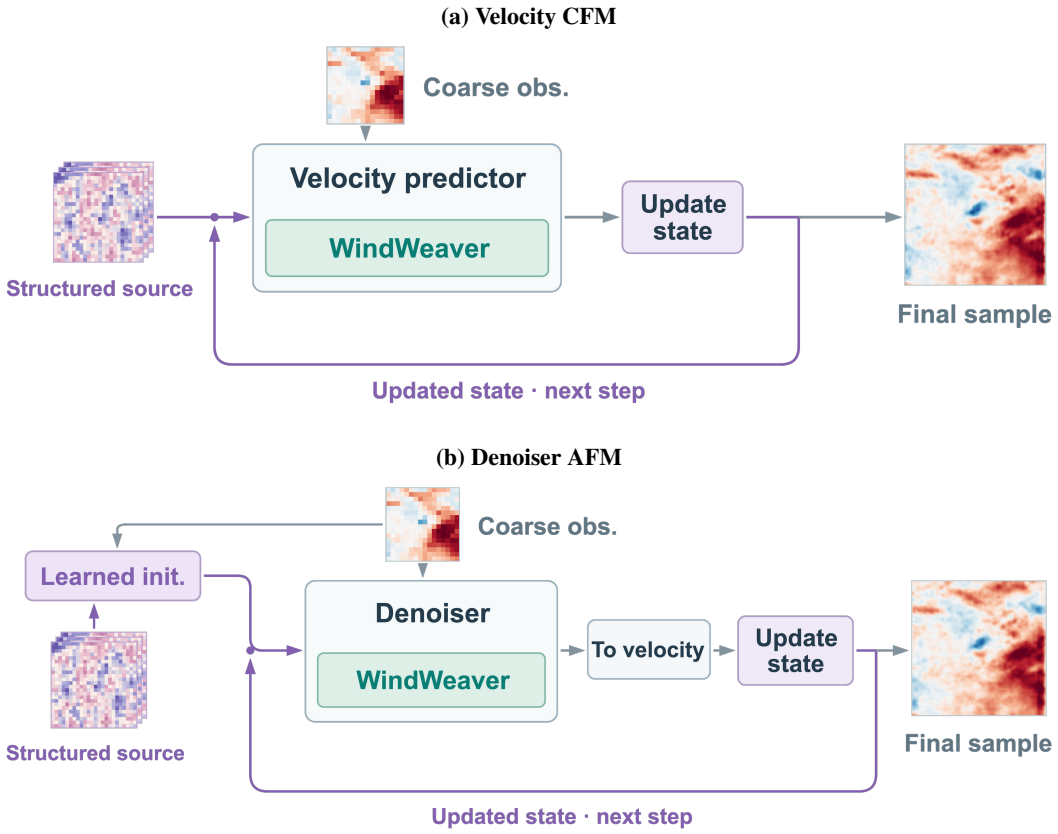

Figure 4: **Initialization and repeated prediction in CFM and AFM.** A structured full-sequence source is sampled once per ensemble member. CFM initializes directly from this source and predicts velocity. AFM combines the source with a learned conditional center and adaptive scale, and converts its denoiser output to velocity. At each Euler step, WindWeaver reads the updated candidate's latent features under fixed coarse conditions.

B IMPLEMENTATION AND ALGORITHMS

B.1 OBSERVATION OPERATORS FOR THE THREE TASKS

Observation constraints follow the input construction. All methods preserve supplied aggregates in the controlled East Asian ERA5 and same-product CERRA tasks. ERA5→CERRA pairs distinct products and permits corrections to coarse means. The spatial operator $\mathcal{Q}$ separates prediction scales, while $\mathcal{P}_H$ constrains residual aggregates according to each task.

Spatial aggregation. For a block B_j and positive cell-area weights a_p, define

$$(Dx)_{c,k,j} = \frac{\sum_{p\in B_j} a_p x_{c,k,p}}{\sum_{p\in B_j} a_p}, \qquad [U(m)]_{c,k,p} = m_{c,k,j} \quad (p \in B_j). \tag{7}$$

East Asian ERA5 uses 6×6 blocks and $a_p = \cos\varphi_p$ on its $0.25°$ grid, giving $1.5°$ coarse cells. Same-product CERRA uses 4×4 area-weighted pooling of native 5.5 km cells, giving 22 km inputs. Patch origins are multiples of four, keeping local pooling aligned with the global grid. ERA5→CERRA uses 6×6 target blocks for the readout decomposition and interpolates native ERA5 conditions onto aligned 33 km centers. Both European tasks use the cached Lambert-grid cell areas.

Instantaneous and interval observations. Let k index frames, with spacing $\Delta t = 1\,\mathrm{h}$ for East Asian ERA5 and 3 h for CERRA. Put $L = 6\,\mathrm{h}/\Delta t$, so $k = 0,\dots,2L$. Channels $c = 1,\dots,8$ are instantaneous; radiation channel 9 at frame k is the mean flux over the preceding Δt interval. For the controlled tasks, the observation vector is

$$\mathcal{H}X = \left(\{DX_{c,k} : c \le 8,\ k \in \{0, L, 2L\}\}, \quad \frac{1}{L}\sum_{k=1}^{L} DX_{9,k}, \quad \frac{1}{L}\sum_{k=L+1}^{2L} DX_{9,k} \right). \tag{8}$$

For East Asian ERA5 this constrains instantaneous frames 0/6/12 and radiation means over frames 1–6 and 7–12. For same-product CERRA it constrains instantaneous frames 0/2/4 and radiation pairs 1–2 and 3–4, corresponding to hours 3/6 and 9/12. The remaining instantaneous block means are free. Radiation at frame zero has a preceding partner or interval outside the generated window and is excluded from these constraints and the primary radiation score. Coarse inputs at -12 and -6 h, and the six-hour radiation mean ending at zero, remain conditioning context.

The right inverse R_H replicates each instantaneous observation over its spatial block at the observed frame. For a radiation observation it repeats the value over the block and all L frames of its interval; unspecified entries are zero. Consequently $\mathcal{H}R_H = I$ and

$$\mathcal{P}_H = I - R_H\mathcal{H}, \qquad \mathcal{H}\mathcal{P}_H = 0. \tag{9}$$

Operationally, this subtracts instantaneous residual block means at observed frames and a common residual block–interval mean from each constrained radiation interval. The same operation is applied to residual targets, source draws, predictor outputs, and Euler states. For ERA5→CERRA, $\mathcal{P}_H = I$ at all these locations.

Anchor construction. Instantaneous coarse fields are linearly interpolated between hours 0, 6, and 12 and replicated over spatial blocks. For a constrained radiation interval I, coarse mean $y_{9,I,j}$, and nonnegative deterministic solar proxy $q_{k,p}$, define

$$\bar{q}_{I,j} = \frac{1}{|I|}\sum_{k\in I}(Dq_k)_j, \qquad a_{9,k,p} = y_{9,I,j}\begin{cases} q_{k,p}/\bar{q}_{I,j}, & \bar{q}_{I,j} > 10^{-5}, \\ 1, & \text{otherwise}, \end{cases} \quad p \in B_j. \tag{10}$$

Both branches preserve the supplied interval mean. At frame zero, the anchor replicates the preceding coarse six-hour radiation mean. The anchor allocates radiation using conditions and solar geometry; R_H applies uniform corrections. The paired-product task uses analogous anchor construction and learns residual corrections to target-product means. Cached validity flags and finite-value checks reject incomplete windows, keeping all block cells and their weights intact.

B.2 SCALE DECOMPOSITION AND OBSERVATION CONSISTENCY

The identities $DU = I$ and $\mathcal{Q} = I - UD$ imply

$$D\mathcal{Q} = 0, \quad \mathcal{Q}U = 0, \quad \mathcal{Q}^2 = \mathcal{Q}, \qquad x = U(Dx) + \mathcal{Q}(x). \tag{11}$$

Hence, for $g = U(m) + \mathcal{Q}(d)$, $Dg = m$ and $\mathcal{Q}(g) = \mathcal{Q}(d)$. Any field can be represented by taking $m = Dx$ and $d = x$. The split assigns each readout a spatial component while retaining full field expressivity. Under the area-weighted spatial inner product, UD and $\mathcal{Q}$ are complementary orthogonal projections, giving

$$\|U(\delta m) + \mathcal{Q}(\delta d)\|_A^2 = \|U(\delta m)\|_A^2 + \|\mathcal{Q}(\delta d)\|_A^2. \tag{12}$$

This identity describes the output geometry; the branches share learned features and parameters.

For each controlled task, the uniform R_H also makes $\mathcal{P}_H$ orthogonal in the area-weighted space–time inner product: it removes constants over each constrained block or block–interval. The channel-constant scales S commute with these within-channel operations. Since $\mathcal{H}a = y$ and $\mathcal{H}z_1 = 0$,

$$\mathcal{H}\widehat{X} = \mathcal{H}(a + Sz_1) = y. \tag{13}$$

The specified observed aggregates are preserved in exact arithmetic, while missing instantaneous block means remain learnable. Cloud values outside $[0, 1]$ and negative radiation remain possible; primary evaluation uses raw samples. Paired-product predictions can adjust ERA5 means to match CERRA.

B.3 WIND COORDINATES, FEATURE MIXING, AND ARCHITECTURE

The predictor concatenates nine candidate channels with 64 condition channels. Conditions include the nine-channel anchor and all five coarse observations, giving 45 channels repeated at each target time. The remaining conditions are latitude, longitude, elevation, land–sea mask, four calendar channels, one solar proxy, and one observed-time indicator. Calendar channels are sine/cosine pairs for time of day and day of year. The stem is a $73 \to 96$ convolution with a 3×3 kernel and SiLU activation. Two 3×3 stride-two convolutions encode a quarter-resolution representation of width 192.

Spatial and weather-time sinusoidal embeddings are added to the representation. Each of four blocks contains pre-normalized spatial self-attention within each weather time, temporal self-attention at each latent location, and a feed-forward layer of hidden width $4d$. Blocks use six attention heads, residual connections, and no dropout; generation-time embeddings condition every block. All compared flow variants retain these spatial and temporal attention operations.

Wind communication follows the final attention block. A 1×1 convolution projects the width-192 representation into three groups of 64 channels. Previous, current, and next features are concatenated and passed through a $576 \to 192 \to 192$ pointwise mixer, with GroupNorm(8), SiLU, and a zero-initialized final convolution. The result is added to the unprojected current representation. The mean branch bilinearly resizes these features to the coarse block grid and applies four 1×1 heads for the variable groups. The detail branch resizes to the fine grid, concatenates the width-96 stem skip, and applies a 3×3 decoder and four 3×3 heads. Prediction heads are zero-initialized. AFM's center encoder uses the direct readout, width 96, depth two, and the same task-specific final observation projection.

Coordinate units and boundaries. East Asian ERA5 converts east/north wind into fine-grid displacement per hour as $\big(3600u/[111195 \cdot 0.25 \cos\varphi],\ -3600v/[111195 \cdot 0.25]\big)$. The negative meridional sign follows the southward row coordinate. CERRA uses east/north wind, the cached inverse Lambert Jacobian, 5,500 m fine-grid spacing, and 10,800 seconds per target step. Displacements from the two neighboring frames are averaged at the target location. Feature coordinates account for the actual encoder stride of four. Bilinear feature sampling uses aligned corners; a coordinate outside the latent support falls back to the current-time projected feature for that group. At the temporal boundary, the missing neighbor is replaced by the endpoint itself. Geometry is derived from the coarse-input anchor and cached for the member's sequence; every flow evaluation samples newly computed candidate features at those coordinates.

Static and Reverse retain the learned projection and mixer with zero and negated displacement, respectively. Static therefore exchanges features at a fixed location; No-exchange removes the communication branch. Reversing coordinates also changes boundary fallback locations, an implementation effect included in this comparison. The capacity control adds a $192 \to 485 \to 192$ pointwise MLP after the final attention block, with SiLU and a zero-initialized final linear layer.

B.4 Structured source and normalization

The source is constructed in normalized residual coordinates. The spatial operator K is reflection-padded 3×3 average pooling multiplied by three. Its temporal filter is

$$(T_\rho K\eta)_k = \frac{\sum_{j=0}^{J_k} \rho^j (K\eta)_{k-j}}{\sqrt{\sum_{j=0}^{J_k} \rho^{2j}}}, \qquad \rho = e^{-\Delta t/(3\,\mathrm{h})}, \qquad J_k = \min\left(k, \left\lfloor \frac{3\,\mathrm{h}}{\Delta t} \right\rfloor\right). \tag{14}$$

The unnormalized weights are $[1, e^{-1/3}, e^{-2/3}, e^{-1}]$ for hourly targets and $[1, e^{-1}]$ for three-hourly targets. At the initial sequence boundary the filter is truncated and its squared-weight normalizer is recomputed. Independent white-noise tensors are drawn across all channels, frames, and spatial locations.

For $b = \mathcal{P}_H \epsilon$ and $r = \mathcal{P}_H(\epsilon/2 + \sqrt{3}\, T_\rho K\eta/2)$, the normalized source is

$$\xi_c = r_c \sqrt{\frac{\langle b_c^2 \rangle}{\max(\langle r_c^2 \rangle, 10^{-12})}}. \tag{15}$$

Brackets average all frames and spatial cells uniformly within one sample/channel. Whole-window scaling preserves the observation nullspace and matches the sample norm in normalized residual coordinates. The draw-dependent gain means the source is not an exact fixed-covariance Gaussian. Projection and norm matching can also couple initially independent random inputs.

The no-memory control removes T_ρ; the no-space control replaces K by the identity. Both retain the independent tensors, mixture weights, projection, and gain. The independent-source control uses $\mathcal{P}_H \epsilon$. In controlled tasks, radiation projection couples frames within a constrained interval even after temporal filtering is removed. Each ensemble member receives one full-sequence source draw and follows deterministic Euler updates. Source correlation stays on the fixed grid; wind guides feature sampling.

Spatial variance at the boundary. Before projection and norm matching, $K\eta$ has variance 1, $5/3$, and $25/9$ at interior, edge, and corner cells, respectively, for unit-variance independent input. Reflection repeats filter coefficients, yielding corner weights of $4, 2, 2, 1$, divided by three. The corresponding mixed-source variances are 1, $3/2$, and $7/3$ before subsequent operations. Squared-weight temporal normalization preserves these factors; spatial variation remains after whole-window normalization. The no-space comparison therefore changes both local variance structure and spatial correlation.

B.5 Host objectives and AFM closure

For a nonnegative spatial weight $w_{b,p}$ and groups $\mathcal{G}$ containing six wind components, temperature, cloud, and radiation, define

$$\|e\|_{\mathcal{G},A}^2 = \frac{1}{4} \sum_{g \in \mathcal{G}} \frac{\sum_{b,c \in g,k,p} w_{b,p} e_{b,c,k,p}^2}{|g| T \sum_{b,p} w_{b,p}}. \tag{16}$$

Weights combine cell area with the task's scoring ownership or central crop. Training includes every target frame. The loss is computed after residual normalization, so each variable group receives equal weight in the objective.

For AFM, put $u = 1 - s$, $\sigma = u\sigma_z$, and $\sigma_{\mathrm{data}} = 1$. The EDM-preconditioned denoiser is

$$\begin{aligned} \mathcal{D}(z_s, \sigma; Y) &= c_{\mathrm{skip}} z_s + c_{\mathrm{out}} f_{\theta,\phi}(c_{\mathrm{in}} z_s, c_{\mathrm{noise}}; Y), \\ c_{\mathrm{skip}} &= (1+\sigma^2)^{-1}, \quad c_{\mathrm{out}} = \frac{\sigma}{\sqrt{1+\sigma^2}}, \quad c_{\mathrm{in}} = (1+\sigma^2)^{-1/2}, \quad c_{\mathrm{noise}} = \tfrac{1}{4}\log\sigma. \end{aligned} \tag{17}$$

The implementation passes $\log \sigma/4000$ to the time interface because the backbone multiplies that argument by 1,000. With the initial states and straight paths defined in Section 3.5, the objectives are

$$\mathcal{L}_{\mathrm{CFM}} = \mathbb{E}\|\widehat{v} - (z_\star - z_0)\|^2_{\mathcal{G},A},$$
$$\mathcal{L}_{\mathrm{AFM}} = \mathbb{E}\left\|\frac{\mathcal{D} - z_\star}{1-s}\right\|^2_{\mathcal{G},A} + \frac{1}{4}\mathbb{E}\left\|\frac{\bar{\mu}_\psi(Y) - z_\star}{\sigma_z}\right\|^2_{\mathcal{G},A}. \tag{18}$$

The equality $\widehat{v} - (z_\star - z_0) = (\mathcal{D} - z_\star)/(1-s)$ follows from the straight path, including with the structured source. Training samples s uniformly from $[0, 1)$ and floors $1 - s$ at 10^{-6}. The AFM center receives gradients from both terms; σ_z is detached during gradient updates and calibrated to the task-specific source distribution.

Center projection and scale calibration. The center encoder returns $\bar{\mu}_\psi = \mathcal{P}_H \mu_\psi$. In the controlled tasks, $\mathcal{H}\bar{\mu}_\psi = \mathcal{H}\xi = \mathcal{H}z_\star = 0$, so the initial state and every point on the training path lie in the observation nullspace. The skip and output coefficients in Eq. (17) are sample-wise scalars, preserving this subspace; the predictor also ends with $\mathcal{P}_H$. The same projected center is used in initialization, training, regularization, calibration, and sampling.

Calibration uses 64 fixed windows from the training years. Let r_k be the square root of their mean group-balanced center MSE in normalized residual coordinates. Initialization sets $\sigma_z = \max(0.002, r_0)$; every 1,000 updates sets $\sigma_z \leftarrow \max(0.002, 0.9\sigma_z + 0.1r_k)$. Calibration evaluates the current model with gradients disabled and copies the updated buffer to the exponential moving average (EMA) model. Saved checkpoints retain the scale buffer.

B.6 Training and sampling algorithms

Algorithms 1 and 2 expose the shared interfaces. The predictor operation includes the backbone, repeated wind-guided feature retrieval, learned mixing, and task-specific readout. Its inputs are the raw state and generation time for CFM, and the preconditioned state and noise embedding for AFM.

Algorithm 1 Joint training of WindWeaver and its flow host

Require: Paired training windows (Y, X), fixed scales S, task projector $\mathcal{P}_H$, host choice
1: Initialize predictor parameters and, for AFM, center encoder and calibrated σ_z
2: **for** each optimization update **do**
3: Build $a(Y)$, conditioning features, and wind coordinates $G(Y)$
4: $z_\star \leftarrow \mathcal{P}_H S^{-1}(X - a(Y))$; draw independent ϵ, η
5: $\xi \leftarrow \mathrm{Source}(\epsilon, \eta, \mathcal{P}_H)$ using Eqs. (14)–(15)
6: $z_0 \leftarrow \xi$ for CFM; $z_0 \leftarrow \bar{\mu}_\psi(Y) + \sigma_z \xi$ for AFM
7: Draw $s \sim \mathcal{U}[0, 1)$; set $z_s \leftarrow (1-s)z_0 + sz_\star$
8: Evaluate current-candidate features with $G(Y)$ and form $\widehat{v}$ or $\mathcal{D}$
9: Jointly update trainable parameters with Eq. (18); update parameter EMA
10: **if** AFM and the update is a multiple of 1,000 **then**
11: Recalibrate σ_z on the fixed training bank and copy its buffer to EMA
12: **end if**
13: **end for**
14: Select the EMA checkpoint with lowest validation balanced CRPS

B.7 Data preparation and training protocol

Spatial and temporal sampling. East Asian ERA5 uses a 144×264 rectangular grid with centers at 18.25–54°N and 72–137.75°E. Six overlapping 96×96 contexts supply predictions, with unique area-weighted scoring ownership and a longitude mask of 73–135°E. The scored region includes neighboring land and ocean areas and uses no administrative-boundary mask. European tasks use 192×192 contexts and central 96×96 scoring crops, sampled from 288 eligible origins across six geographic strata. Both evaluations include land and ocean.

All three tasks condition on times $-12, -6, 0, 6, 12$ h. East Asian ERA5 targets are hourly from 0 to 12 h; CERRA targets are at 0/3/6/9/12 h. Same-product CERRA conditions use only pooled CERRA

Algorithm 2 Generate one ensemble member with the selected EMA model

Require: Conditions Y, fixed scales S, projector $\mathcal{P}_H$, number of Euler steps N
Build $a(Y)$, conditioning features, and cached wind coordinates $G(Y)$
Draw independent ϵ, η once for the full sequence; construct ξ
$z \leftarrow \xi$ for CFM; $z \leftarrow \bar{\mu}_\psi(Y) + \sigma_z \xi$ for AFM
for $k = 0, \ldots, N-1$ **do**
    $s \leftarrow k/N$
    if CFM **then**
        $\widehat{v} \leftarrow f_{\theta,\phi}(z, s; Y, G)$
    else
        $\sigma \leftarrow (1-s)\sigma_z$; $\widehat{v} \leftarrow [\mathcal{D}(z, \sigma; Y, G) - z]/(1-s)$
    end if
    $z \leftarrow \mathcal{P}_H[z + \widehat{v}/N]$
end for
return $\widehat{X} = a(Y) + Sz$

fields, including historical fields outside the target window. ERA5→CERRA conditions use ERA5 dynamic fields. The six pressure-level wind components use a common east/north convention. Static features are elevation and land–sea mask: the East Asian mask is an all-zero historical placeholder, while European masks contain valid product data. All methods within a task receive the same static features. Pressure-level values below terrain are retained as reanalysis-product values.

Radiation and units. Radiation is indexed by interval end time. ERA5 hourly accumulated energy is divided by 3,600 to obtain the mean flux over $(e-1, e]$. A six-hour input ending at e sums the six hourly records ending at $e-5, \ldots, e$ and divides by 21,600. A CERRA target is the short-forecast accumulation with start step 0 and end step 3, divided by 10,800 and indexed by validity time e, giving support $(e-3, e]$. The CERRA preparation checks these start and end steps. Same-product CERRA inputs average the two successive three-hour mean fluxes ending at $e-3$ and e; they consequently share exact interval aggregates with their targets. ERA5→CERRA aligns these six-hour supports across products while allowing their values to differ.

Before residual normalization, wind is divided by 20 m/s and temperature is transformed as $(T - 280\,\mathrm{K})/10\,\mathrm{K}$. Cloud is expressed as a fraction, and radiation is divided by 500 W/m^2. The physical residual scale is $S_{cc} = b_c s_c$, with $b_c = (20, 20, 20, 20, 20, 20, 10, 1, 500)_c$ and s_c the frozen channel residual scale in these preprocessed units. The fixed temperature offset cancels in residuals. Each task estimates s_c as the area-weighted residual RMS on 256 fixed training windows, with a 0.01 floor. Scales are frozen before model fitting and shared among compared variants within each task (Table 4).

Table 4: Frozen residual scales s_c in preprocessed coordinates (rounded to eight decimal places). Multiply by b_c to obtain the physical residual scales in Eq. (1).

Channel	East Asian ERA5	Same-product CERRA	ERA5→CERRA
u_{850}	0.07489181	0.05444681	0.07856146
v_{850}	0.06882316	0.05569449	0.07964113
u_{700}	0.07510684	0.05257910	0.06988771
v_{700}	0.06488177	0.05295898	0.06939224
u_{300}	0.12200847	0.08732963	0.10032405
v_{300}	0.10301565	0.10171457	0.11166748
Temperature	0.24642046	0.09397306	0.14281707
Cloud	0.18112400	0.13558482	0.27633193
Radiation	0.13835807	0.10811622	0.13217357

Splits and optimization. Training uses 2020–2021. In each month of 2022, validation starts on day 2 at 00 UTC, and a reserved development split starts on day 4. Test windows start on day 16 at 00 and 12 UTC. Six contexts or strata per start yield 72 validation windows and 144 test windows

per task. Incomplete year-edge windows are excluded, and cached data-validity masks remain fixed. Overlapping contexts and adjacent test windows share weather information.

Each run uses 60 epochs of 4,800 sampled training windows, giving 36,000 optimizer updates with microbatch size four and two-step gradient accumulation. CFM uses AdamW with learning rate 2×10^{-4} and weight decay 10^{-4}. AFM uses Adam with learning rate 10^{-4}, $\beta = (0.9, 0.99)$, and zero weight decay. Both use gradient clipping at one, parameter EMA 0.999, 200-update warmup, and cosine learning-rate decay to 10% of the initial rate. The backbone uses bfloat16 autocasting on CUDA; sampled features are interpolated in float32. Seeds are 42, 43, and 44.

Checkpoint selection and generation. Every five epochs, the EMA model is evaluated with four members and 20 Euler steps. Validation balanced CRPS selects one checkpoint per run, reused across variables and diagnostics. The standard test setting uses 12 members and 30 Euler steps. The two independent source tensors use separate random generators; matched variants use the same case/member seeds. Source sampling and the AFM center are evaluated once per member, and the flow predictor once per Euler step. Feature retrieval and AFM center computation also contribute to wall-clock cost.

B.8 Scoring definitions and aggregation

Empirical ensemble CRPS. For scalar predictions $x^{(1)}, \ldots, x^{(M)}$ and reference y, the implemented continuous ranked probability score (CRPS) is

$$C_M(x, y) = \frac{1}{M} \sum_{m=1}^{M} |x^{(m)} - y| - \frac{1}{2M^2} \sum_{m=1}^{M} \sum_{n=1}^{M} |x^{(m)} - x^{(n)}|. \tag{19}$$

Equation (19) scores the empirical ensemble distribution (Gneiting & Raftery, 2007), with $M = 12$ throughout testing. Fair ensemble CRPS uses $2M(M-1)$ in the pair-term denominator (Ferro, 2014). For independent draws from an underlying distribution F, the expectation of Eq. (19) is $\mathrm{CRPS}(F, y) + \mathbb{E}|X - X'|/(2M)$. Fixing ensemble size standardizes the comparison, while the finite-ensemble term depends on forecast spread.

Physical scores and masks. Let b index evaluation windows and write $w_{g,b,k,p} = a_{b,p} o_{b,p} r_{g,b,k,p}$. Here a is cell area and r is the variable group's time/daylight mask. The factor o assigns unique spatial ownership in East Asian ERA5 and selects central crops in the European tasks. Wind, temperature, and cloud primary scores cover missing target times: hours 1–5 and 7–11 in East Asia, or 3 and 9 h for CERRA. Radiation includes frames $k \geq 1$ for which the solar proxy exceeds 0.02. The proxy uses the interval midpoint for hourly targets and averages three hourly midpoints for CERRA three-hour targets. Both land and ocean contribute, and primary scores use raw, unclipped predictions.

For physical-unit pointwise CRPS $C_{b,c,k,p}$ and ensemble mean $\bar{x}$, the group scores within one training seed are

$$\begin{aligned} \mathrm{CRPS}_g &= \frac{\sum_{b,c\in g,k,p} w_{g,b,k,p} C_{b,c,k,p}}{|g| \sum_{b,k,p} w_{g,b,k,p}}, \\ \mathrm{RMSE}_g &= \left(\frac{\sum_{b,c\in g,k,p} w_{g,b,k,p} (\bar{x}_{b,c,k,p} - y_{b,c,k,p})^2}{|g| \sum_{b,k,p} w_{g,b,k,p}} \right)^{1/2}. \end{aligned} \tag{20}$$

Numerators and denominators are accumulated across all windows before division, and RMSE takes the square root after this aggregation. Wind pools six components in these sums; each other group has one channel. Reported means and sample standard deviations are then computed over the three training seeds. The reanalysis target is the scoring reference.

Balanced checkpoint criterion. Channel residual scales normalize CRPS before aggregation:

$$B = \frac{1}{4} \sum_{g\in\mathcal{G}} \frac{\sum_{b,c\in g,k,p} w_{g,b,k,p} C_{b,c,k,p}/(b_c s_c)}{|g| \sum_{b,k,p} w_{g,b,k,p}}. \tag{21}$$

Here b_c is the channel preprocessing divisor and s_c is the training residual scale, so $b_c s_c = S_{cc}$. The criterion uses the primary masks and spatial weights above, with zero-weight entries excluded from both sums. Each channel is normalized before variable-group averaging, including the six wind channels. Checkpoint selection uses B, with equal group weights on a task-specific normalized scale.

Temporal increments. Pairwise-difference diagnostics probe dependence in multivariate verification (Scheuerer & Hamill, 2015). We evaluate the empirical distribution of signed temporal differences using CRPS and ensemble-mean RMSE. At lag τ, each member predicts the signed difference $x^{(m)}(t+\tau) - x^{(m)}(t)$, scored against the corresponding reference difference. Member identities and physical variable units are preserved; increments are unscaled by τ. Evaluated lags are 1/2/3/6 h in East Asia and 3/6/9 h for CERRA. Instantaneous-variable pairs include at least one missing endpoint; the strict subset has both endpoints missing. In CERRA, the strict six-hour subset is 3→9 h, matching the main six-hour instantaneous comparison. Radiation pairs compare valid interval-mean fluxes at frames after zero, with solar proxy above 0.02 at both endpoints. Pair counts and valid supports can differ across lags.

Temporal uncertainty intervals use 2,000 paired resamples of the 12 month blocks, keeping Full and control conditions matched and the three trained seeds fixed. Each draw reaggregates per-case weighted numerators and denominators before calculating paired improvement. The 2.5th and 97.5th percentiles define the 95% interval for month-block variation; sample SD summarizes training-seed variation.

C Additional reconstruction and temporal results

C.1 Complete component comparisons and score units

Tables 5–10 compare all 18 configurations for each protocol over three training seeds, with sample SD where available. Balanced CRPS averages the four variable groups after normalization by protocol-specific training residual scales. Same-product CERRA reports normalized variable CRPS; the other two protocols report physical-unit CRPS, and all RMSE values are in physical units. Table 1 expresses these comparisons as relative reductions. Rounding of source means can produce small differences in the reported percentages.

No-exchange removes the wind-guided branch and retains backbone attention. Static uses zero displacement within this branch; Reverse flips the wind-derived displacement. The source-memory and source-smoothing controls remove temporal filtering and local spatial filtering, respectively. Removing spatial filtering also changes local variance, particularly at reflected boundaries. Each control measures a component's contribution conditional on the remaining design. Full has the best balanced CFM score in every protocol, with variable-specific exceptions: ERA5→CERRA CFM without source memory, for example, has slightly lower T2m CRPS.

Table 5: **East Asian ERA5: all state-reconstruction CRPS results.** Variable scores are in physical units.

Configuration	Wind	T2m	Cloud	Radiation	Balanced ± SD
CFM Base	0.627554	0.522063	0.065925	18.127810	0.304177 ± 0.000225
CFM WindWeaver	0.557096	0.471783	0.063445	17.525384	0.283006 ± 0.000235
CFM Scale readouts only	0.616044	0.501558	0.065755	18.038554	0.299341 ± 0.000117
CFM w/o scale readouts	0.573200	0.486824	0.063994	17.703573	0.288550 ± 0.000308
CFM w/o structured source	0.593896	0.498870	0.064719	17.913378	0.294350 ± 0.000115
CFM No-exchange	0.569935	0.474824	0.064193	17.606988	0.286373 ± 0.000291
CFM Capacity control	0.626023	0.521355	0.065787	18.101347	0.303609 ± 0.000281
AFM Base	0.700804	0.569976	0.070225	19.065481	0.329164 ± 0.000315
AFM WindWeaver	0.595401	0.503692	0.065750	18.189193	0.297729 ± 0.000506
AFM Scale readouts only	0.669105	0.544603	0.069390	18.844145	0.319476 ± 0.000187
AFM w/o scale readouts	0.632366	0.527814	0.066924	18.470918	0.308476 ± 0.000070
AFM w/o structured source	0.649959	0.542270	0.068410	18.776239	0.315233 ± 0.000313
AFM No-exchange	0.616440	0.506591	0.066608	18.247154	0.302038 ± 0.000228
AFM Capacity control	0.699533	0.569479	0.070100	19.029155	0.328618 ± 0.000186
CFM Static	0.625799	0.484699	0.066616	17.597466	0.297928 ± 0.000771
CFM Reverse	0.612090	0.510389	0.065024	17.880828	0.298315 ± 0.000557
CFM w/o source memory	0.609397	0.493759	0.064297	17.776808	0.290601 ± 0.000674
CFM w/o source smoothing	0.598792	0.498438	0.064904	17.959132	0.294700 ± 0.000458

Balanced CRPS averages the four groups after normalization by their training residual scales.

Table 6: **East Asian ERA5: ensemble-mean RMSE in physical units.** Means over three training seeds.

Configuration	Wind (m/s)	T2m (K)	Cloud	Radiation (W/m^2)
CFM Base	1.1768	1.0181	0.1513	40.9943
CFM WindWeaver	1.0507	0.9305	0.1463	39.9365
CFM Scale readouts only	1.1650	0.9817	0.1505	40.7821
CFM w/o scale readouts	1.0749	0.9589	0.1474	40.2503
CFM w/o structured source	1.1142	0.9764	0.1485	40.5499
CFM No-exchange	1.0778	0.9362	0.1478	40.0760
CFM Capacity control	1.1734	1.0161	0.1511	40.9513
AFM Base	1.3020	1.0916	0.1573	42.2312
AFM WindWeaver	1.1160	0.9879	0.1501	41.0638
AFM Scale readouts only	1.2548	1.0499	0.1563	41.9277
AFM w/o scale readouts	1.1770	1.0307	0.1514	41.5508
AFM w/o structured source	1.2082	1.0464	0.1544	41.8327
AFM No-exchange	1.1647	0.9931	0.1518	41.1531
AFM Capacity control	1.3000	1.0913	0.1571	42.1925
CFM Static	1.1774	0.9655	0.1530	40.1483
CFM Reverse	1.1566	1.0045	0.1497	40.6397
CFM w/o source memory	1.1289	0.9728	0.1501	40.4583
CFM w/o source smoothing	1.0828	0.9586	0.1498	40.7774

Table 7: **Same-product CERRA: all state-reconstruction CRPS results.** The four variable columns report normalized CRPS.

Configuration	Wind	T2m	Cloud	Radiation	Balanced ± SD
CFM Base	0.7854	0.4218	0.5832	0.1693	0.4899 ± 0.0009
CFM WindWeaver	0.6580	0.3973	0.5267	0.1569	0.4347 ± 0.0008
CFM Scale readouts only	0.7814	0.4174	0.5773	0.1680	0.4860 ± 0.0034
CFM w/o scale readouts	0.6662	0.4025	0.5321	0.1591	0.4400 ± 0.0010
CFM w/o structured source	0.6813	0.4054	0.5412	0.1605	0.4471 ± 0.0009
CFM No-exchange	0.7200	0.4102	0.5511	0.1634	0.4612 ± 0.0115
CFM Capacity control	0.7814	0.4213	0.5821	0.1693	0.4885 ± 0.0023
AFM Base	0.8485	0.4649	0.6239	0.1830	0.5301 ± 0.0021
AFM WindWeaver	0.6821	0.4127	0.5524	0.1646	0.4530 ± 0.0003
AFM Scale readouts only	0.8489	0.4556	0.6247	0.1762	0.5264 ± 0.0019
AFM w/o scale readouts	0.6906	0.4230	0.5615	0.1691	0.4611 ± 0.0018
AFM w/o structured source	0.7197	0.4412	0.5741	0.1694	0.4761 ± 0.0014
AFM No-exchange	0.7772	0.4267	0.5891	0.1696	0.4907 ± 0.0024
AFM Capacity control	0.8495	0.4640	0.6260	0.1837	0.5308 ± 0.0014
CFM Static	0.7453	0.4160	0.5607	0.1642	0.4715 ± 0.0009
CFM Reverse	0.7341	0.4278	0.5801	0.1673	0.4773 ± 0.0011
CFM w/o source memory	0.7249	0.4184	0.5628	0.1609	0.4667 ± 0.0009
CFM w/o source smoothing	0.6667	0.4101	0.5327	0.1620	0.4429 ± 0.0007

Balanced CRPS is the equal mean of the four normalized variable-group scores, before rounding.

Table 8: **Same-product CERRA: ensemble-mean RMSE in physical units.** Means over three training seeds.

Configuration	Wind (m/s)	T2m (K)	Cloud	Radiation (W/m^2)
CFM Base	1.2530	0.5404	0.1172	17.0212
CFM WindWeaver	1.0677	0.5177	0.1094	16.0800
CFM Scale readouts only	1.2465	0.5316	0.1164	16.9944
CFM w/o scale readouts	1.0816	0.5238	0.1098	16.2272
CFM w/o structured source	1.0991	0.5221	0.1102	16.2455
CFM No-exchange	1.1506	0.5285	0.1133	16.6549
CFM Capacity control	1.2453	0.5399	0.1171	17.0135
AFM Base	1.3656	0.5811	0.1231	17.6822
AFM WindWeaver	1.1092	0.5313	0.1128	16.5207
AFM Scale readouts only	1.3775	0.5642	0.1230	17.1691
AFM w/o scale readouts	1.1169	0.5500	0.1142	16.7751
AFM w/o structured source	1.1563	0.5530	0.1144	16.6332
AFM No-exchange	1.2534	0.5415	0.1191	16.9756
AFM Capacity control	1.3662	0.5796	0.1233	17.7501
CFM Static	1.1245	0.5300	0.1121	16.7685
CFM Reverse	1.1670	0.5443	0.1135	17.0032
CFM w/o source memory	1.1058	0.5367	0.1144	16.4935
CFM w/o source smoothing	1.1139	0.5251	0.1106	16.2984

Table 9: **ERA5 → CERRA: all state-reconstruction CRPS results.** Variable scores are in physical units.

Configuration	Wind	T2m	Cloud	Radiation	Balanced ± SD
CFM Base	1.255528	0.608151	0.115014	25.031880	0.488795 ± 0.001346
CFM WindWeaver	1.019379	0.577485	0.104060	22.564624	0.430893 ± 0.000267
CFM Scale readouts only	1.233683	0.606031	0.114193	24.380035	0.481778 ± 0.001328
CFM w/o scale readouts	1.142008	0.577824	0.105649	22.971111	0.449353 ± 0.001090
CFM w/o structured source	1.173785	0.601601	0.109243	23.963384	0.465280 ± 0.002363
CFM No-exchange	1.161216	0.583063	0.108018	22.977362	0.457027 ± 0.000631
CFM Capacity control	1.254090	0.613872	0.115023	25.059762	0.489678 ± 0.000543
AFM Base	1.283034	0.660332	0.116258	24.579675	0.501337 ± 0.000532
AFM WindWeaver	1.212566	0.596065	0.108973	23.214458	0.467909 ± 0.001546
AFM Scale readouts only	1.288789	0.654508	0.116529	24.600340	0.501595 ± 0.007675
AFM w/o scale readouts	1.254723	0.614474	0.108102	23.115035	0.476495 ± 0.003285
AFM w/o structured source	1.280701	0.646766	0.112568	24.385881	0.494705 ± 0.003440
AFM No-exchange	1.233519	0.611924	0.113128	23.484114	0.478696 ± 0.002048
AFM Capacity control	1.285532	0.663587	0.117062	24.751094	0.503604 ± 0.001259
CFM Static	1.222278	0.584976	0.109661	23.755020	0.465431 ± 0.001142
CFM Reverse	1.247216	0.597412	0.110050	23.897065	0.468207 ± 0.000617
CFM w/o source memory	1.044598	0.575232	0.106638	23.119991	0.449006 ± 0.000590
CFM w/o source smoothing	1.122805	0.598026	0.107454	22.684817	0.460586 ± 0.000821

Balanced CRPS averages the four groups after normalization by their training residual scales.

Table 10: **ERA5 → CERRA: ensemble-mean RMSE in physical units.** Means over three training seeds.

Configuration	Wind (m/s)	T2m (K)	Cloud	Radiation (W/m^2)
CFM Base	2.3240	1.1239	0.2424	55.2990
CFM WindWeaver	1.8884	1.0856	0.2259	51.9450
CFM Scale readouts only	2.3034	1.1264	0.2412	54.6005
CFM w/o scale readouts	2.1676	1.0803	0.2288	52.6613
CFM w/o structured source	2.1993	1.1212	0.2319	53.4995
CFM No-exchange	2.1738	1.0939	0.2335	52.8578
CFM Capacity control	2.3230	1.1319	0.2427	55.3519
AFM Base	2.3802	1.2066	0.2445	54.0522
AFM WindWeaver	2.2872	1.1142	0.2330	52.6162
AFM Scale readouts only	2.3953	1.1988	0.2443	54.2682
AFM w/o scale readouts	2.3584	1.1452	0.2325	52.5143
AFM w/o structured source	2.3831	1.1802	0.2365	53.7406
AFM No-exchange	2.3290	1.1396	0.2407	53.2376
AFM Capacity control	2.3809	1.2119	0.2453	54.4449
CFM Static	2.2299	1.0569	0.2303	50.7913
CFM Reverse	2.2543	1.0899	0.2279	51.1363
CFM w/o source memory	1.9522	1.0534	0.2016	47.1054
CFM w/o source smoothing	1.9114	1.0827	0.2256	49.8509

C.2 DETERMINISTIC INTERPOLATION

Validation balanced CRPS selects the joint choice of spatial interpolation, temporal interpolation and solar allocation. The same-product setting also applies conservative correction. Table 11 reports the selected deterministic baselines.

Table 11: **Validation-selected deterministic interpolation.** Physical-unit CRPS; for a deterministic prediction CRPS equals absolute error.

Protocol	Wind	T2m	Cloud	Radiation	Balanced
East Asian ERA5	1.016974	1.434373	0.116510	39.780371	0.603051
ERA5 → CERRA	1.594677	1.101903	0.193008	46.039871	0.775190

East Asian ERA5: bicubic spatial, linear temporal, solar radiation allocation with conservative correction. ERA5 to CERRA: bilinear spatial, linear temporal, solar allocation.

C.3 NET TEMPORAL CONTRIBUTION OF WIND-GUIDED COMMUNICATION

Tables 12–14 report all 24 Full–No-exchange comparisons for six-hour increments. No-exchange retains the source, scale readouts, inputs and attention backbone, isolating the added communication branch. Evaluation uses the original validation-selected EMA checkpoints, 12 members and NFE30 on matched cases and masks. Seed SD summarizes training variation; paired month-block intervals summarize uncertainty in relative improvement conditional on the three fitted models.

Full improves every reported CRPS and ensemble-mean RMSE over No-exchange. The Full $<$ No-exchange $<$ Static $<$ Reverse ordering holds for balanced state CRPS and all twelve variable-wise increment CRPS comparisons. RMSE rankings vary: same-product CERRA cloud increment RMSE favors Full (0.206239), followed by Static (0.212063) and No-exchange (0.213660).

Table 12: **Same-product CERRA: six-hour Full–No-exchange comparison.** CFM increment scores: mean ± sample SD over three seeds. Positive reductions favor Full.

Variable (unit)	Metric	Full	No-exchange	Gain (%)	95% CI (%)
Wind (m/s)	CRPS	1.245099 ± 0.004897	1.430892 ± 0.015113	+12.984	[+10.780, +14.510]
T2m (K)	CRPS	0.528364 ± 0.002179	0.541504 ± 0.002276	+2.427	[+1.862, +2.994]
Cloud (fraction)	CRPS	0.114486 ± 0.000222	0.119002 ± 0.000727	+3.795	[+2.875, +4.456]
Radiation (W/m^2)	CRPS	15.959513 ± 0.034336	16.526995 ± 0.036302	+3.434	[+2.742, +4.321]
Wind (m/s)	RMSE	1.756686 ± 0.014111	2.019650 ± 0.031838	+13.020	[+11.160, +14.933]
T2m (K)	RMSE	0.996985 ± 0.003790	1.018389 ± 0.004502	+2.102	[+1.669, +2.650]
Cloud (fraction)	RMSE	0.206239 ± 0.000397	0.213660 ± 0.001232	+3.473	[+2.691, +4.184]
Radiation (W/m^2)	RMSE	34.487829 ± 0.067810	35.575059 ± 0.070585	+3.056	[+2.476, +3.689]

CIs: 2,000 paired month-block resamples with models fixed. Matched cases/masks, 12 members, NFE30. Gains use the displayed six-decimal means.

Table 13: **ERA5 → CERRA: six-hour Full–No-exchange comparison.** CFM increment scores: mean ± sample SD over three seeds. Positive reductions favor Full.

Variable (unit)	Metric	Full	No-exchange	Gain (%)	95% CI (%)
Wind (m/s)	CRPS	1.386716 ± 0.005450	1.607006 ± 0.006250	+13.708	[+12.083, +15.899]
T2m (K)	CRPS	0.651581 ± 0.000868	0.664598 ± 0.001364	+1.959	[+1.568, +2.384]
Cloud (fraction)	CRPS	0.148457 ± 0.000417	0.153638 ± 0.000493	+3.372	[+2.712, +4.164]
Radiation (W/m^2)	CRPS	39.184805 ± 0.297085	39.929766 ± 0.309261	+1.866	[+1.504, +2.363]
Wind (m/s)	RMSE	2.552522 ± 0.014325	3.095439 ± 0.013136	+17.539	[+15.864, +19.778]
T2m (K)	RMSE	1.230742 ± 0.001727	1.249481 ± 0.002763	+1.500	[+1.010, +1.936]
Cloud (fraction)	RMSE	0.296523 ± 0.000841	0.306845 ± 0.001605	+3.364	[+2.529, +3.939]
Radiation (W/m^2)	RMSE	80.359963 ± 0.464655	81.523157 ± 0.418829	+1.427	[+0.974, +1.842]

CIs: 2,000 paired month-block resamples with models fixed. Matched cases/masks, 12 members, NFE30. Gains use the displayed six-decimal means.

Table 14: **East Asian ERA5: six-hour Full–No-exchange comparison.** CFM increment scores: mean ± sample SD over three seeds. Positive reductions favor Full.

Variable (unit)	Metric	Full	No-exchange	Gain (%)	95% CI (%)
Wind (m/s)	CRPS	0.753960 ± 0.001723	0.767534 ± 0.001109	+1.769	[+1.340, +2.184]
T2m (K)	CRPS	0.569170 ± 0.000494	0.570142 ± 0.000612	+0.170	[+0.103, +0.240]
Cloud (fraction)	CRPS	0.097259 ± 0.000109	0.097876 ± 0.000109	+0.630	[+0.395, +0.900]
Radiation (W/m^2)	CRPS	28.627497 ± 0.022475	28.661109 ± 0.019395	+0.117	[+0.066, +0.174]
Wind (m/s)	RMSE	1.422388 ± 0.002933	1.450136 ± 0.002191	+1.913	[+1.464, +2.469]
T2m (K)	RMSE	1.089458 ± 0.000964	1.091238 ± 0.000997	+0.163	[+0.096, +0.220]
Cloud (fraction)	RMSE	0.204511 ± 0.000212	0.205550 ± 0.000171	+0.505	[+0.329, +0.714]
Radiation (W/m^2)	RMSE	58.086677 ± 0.036434	58.144657 ± 0.035951	+0.100	[+0.055, +0.148]

CIs: 2,000 paired month-block resamples with models fixed. Matched cases/masks, 12 members, NFE30. Gains use the displayed six-decimal means.

C.4 Temporal scores over all evaluated lags

Temporal evaluation uses the original validation-selected EMA checkpoints, ensemble size and sampling budget. CRPS scores signed member-wise increments; RMSE scores the difference of ensemble means. Increments retain physical units—m/s, K, cloud fraction and W/m^2—without hourly normalization. Wind CRPS averages the six component scores equally; wind RMSE takes the square root of their pooled squared errors.

Wind, T2m and cloud use pairs with at least one missing endpoint. All eligible six-hour pairs have two missing endpoints: $(1, 7)$, $(2, 8)$, $(3, 9)$, $(4, 10)$ and $(5, 11)$ h for East Asian ERA5, and $(3, 9)$ h for both CERRA-target protocols. The main and strict missing-to-missing six-hour scores therefore coincide and appear once. Radiation scores changes in mean flux between valid daylight intervals. Endpoint pairs and daylight support vary with lag.

Table 15: **East Asian ERA5: temporal-increment CRPS.** Physical-unit means; sample SD across three training seeds appears below each mean.

		CFM				AFM	
Variable	Lag	Base	Full	Static	Reverse	Base	Full
Wind	1h	0.3974 (0.0005)	0.3694 (0.0005)	0.3745 (0.0001)	0.3941 (0.0004)	0.4306 (0.0005)	0.3867 (0.0002)
Wind	2h	0.6106 (0.0003)	0.5650 (0.0010)	0.5757 (0.0002)	0.6021 (0.0005)	0.6581 (0.0008)	0.5913 (0.0003)
Wind	3h	0.7096 (0.0002)	0.6501 (0.0012)	0.6633 (0.0002)	0.6982 (0.0007)	0.7711 (0.0011)	0.6828 (0.0005)
Wind	6h	0.8413 (0.0016)	0.7540 (0.0017)	0.7699 (0.0002)	0.8183 (0.0019)	0.9359 (0.0021)	0.7961 (0.0011)
T2m	1h	0.3172 (0.0032)	0.3021 (0.0014)	0.3026 (0.0012)	0.3147 (0.0011)	0.3346 (0.0012)	0.3140 (0.0016)
T2m	2h	0.4295 (0.0016)	0.4096 (0.0006)	0.4102 (0.0002)	0.4258 (0.0015)	0.4528 (0.0016)	0.4266 (0.0015)
T2m	3h	0.4901 (0.0017)	0.4647 (0.0004)	0.4654 (0.0003)	0.4854 (0.0009)	0.5175 (0.0015)	0.4850 (0.0015)
T2m	6h	0.6100 (0.0029)	0.5692 (0.0005)	0.5704 (0.0005)	0.5999 (0.0015)	0.6447 (0.0016)	0.5977 (0.0030)
Cloud	1h	0.05552 (0.00002)	0.05438 (0.00006)	0.05463 (0.00006)	0.05539 (0.00007)	0.05764 (0.00007)	0.05568 (0.00003)
Cloud	2h	0.07731 (0.00002)	0.07541 (0.00006)	0.07587 (0.00008)	0.07698 (0.00004)	0.08062 (0.00003)	0.07740 (0.00007)
Cloud	3h	0.08686 (0.00002)	0.08439 (0.00008)	0.08494 (0.00008)	0.08647 (0.00009)	0.09093 (0.00004)	0.08680 (0.00011)
Cloud	6h	0.10054 (0.00005)	0.09726 (0.00011)	0.09797 (0.00008)	0.10003 (0.00008)	0.10574 (0.00009)	0.10023 (0.00017)
Radiation	1h	21.998 (0.013)	21.733 (0.024)	21.768 (0.013)	21.973 (0.019)	22.625 (0.018)	22.229 (0.026)
Radiation	2h	28.919 (0.021)	28.334 (0.036)	28.386 (0.009)	28.821 (0.023)	29.941 (0.040)	29.164 (0.031)
Radiation	3h	32.138 (0.035)	31.284 (0.036)	31.354 (0.018)	32.013 (0.056)	33.413 (0.050)	32.303 (0.032)
Radiation	6h	29.469 (0.033)	28.627 (0.022)	28.673 (0.007)	29.350 (0.047)	30.713 (0.058)	29.555 (0.037)

Pairs include a missing endpoint; radiation uses valid daylight intervals. Signed increments retain physical units; eligible pairs vary with lag.

Table 16: **East Asian ERA5: temporal-increment RMSE.** Physical-unit means; sample SD across three training seeds appears below each mean.

		CFM				AFM	
Variable	Lag	Base	Full	Static	Reverse	Base	Full
Wind	1h	0.7911 (0.0004)	0.7405 (0.0006)	0.7541 (0.0002)	0.7805 (0.0005)	0.8424 (0.0005)	0.7687 (0.0006)
Wind	2h	1.1937 (0.0004)	1.1018 (0.0012)	1.1281 (0.0006)	1.1766 (0.0009)	1.2778 (0.0011)	1.1478 (0.0004)
Wind	3h	1.3633 (0.0007)	1.2450 (0.0018)	1.2757 (0.0007)	1.3466 (0.0010)	1.4759 (0.0023)	1.3026 (0.0005)
Wind	6h	1.5861 (0.0043)	1.4224 (0.0029)	1.4569 (0.0006)	1.5547 (0.0033)	1.7601 (0.0060)	1.4934 (0.0015)
T2m	1h	0.6134 (0.0046)	0.5844 (0.0022)	0.5852 (0.0022)	0.6079 (0.0027)	0.6309 (0.0021)	0.6045 (0.0032)
T2m	2h	0.8216 (0.0025)	0.7857 (0.0010)	0.7867 (0.0006)	0.8156 (0.0023)	0.8524 (0.0020)	0.8142 (0.0024)
T2m	3h	0.9327 (0.0030)	0.8870 (0.0007)	0.8883 (0.0006)	0.9267 (0.0018)	0.9716 (0.0019)	0.9219 (0.0024)
T2m	6h	1.1625 (0.0054)	1.0895 (0.0010)	1.0916 (0.0006)	1.1474 (0.0022)	1.2138 (0.0022)	1.1402 (0.0054)
Cloud	1h	0.12751 (0.00006)	0.12533 (0.00004)	0.12587 (0.00006)	0.12713 (0.00008)	0.13018 (0.00011)	0.12713 (0.00004)
Cloud	2h	0.17142 (0.00008)	0.16756 (0.00006)	0.16850 (0.00011)	0.17095 (0.00013)	0.17599 (0.00024)	0.17062 (0.00014)
Cloud	3h	0.18840 (0.00008)	0.18345 (0.00010)	0.18455 (0.00011)	0.18744 (0.00010)	0.19408 (0.00029)	0.18727 (0.00021)
Cloud	6h	0.21071 (0.00005)	0.20451 (0.00021)	0.20581 (0.00010)	0.20976 (0.00019)	0.21793 (0.00032)	0.20924 (0.00029)
Radiation	1h	48.462 (0.029)	47.996 (0.039)	48.059 (0.022)	48.363 (0.043)	49.256 (0.032)	48.773 (0.030)
Radiation	2h	61.312 (0.053)	60.258 (0.058)	60.333 (0.016)	61.135 (0.035)	62.630 (0.068)	61.596 (0.040)
Radiation	3h	66.701 (0.094)	65.088 (0.061)	65.215 (0.042)	66.524 (0.050)	68.327 (0.085)	66.775 (0.063)
Radiation	6h	59.519 (0.045)	58.087 (0.036)	58.171 (0.038)	59.266 (0.048)	61.175 (0.092)	59.599 (0.046)

Pairs include a missing endpoint; radiation uses valid daylight intervals. Signed increments retain physical units; eligible pairs vary with lag.

Table 17: **Same-product CERRA: temporal-increment CRPS.** Physical-unit means; sample SD across three training seeds appears below each mean.

		CFM				AFM	
Variable	Lag	Base	Full	Static	Reverse	Base	Full
Wind	3h	1.0474 (0.0060)	0.8893 (0.0038)	1.0013 (0.0113)	1.0163 (0.0043)	1.1373 (0.0048)	0.9116 (0.0020)
Wind	6h	1.5383 (0.0117)	1.2451 (0.0049)	1.4575 (0.0190)	1.4747 (0.0091)	1.6773 (0.0118)	1.2713 (0.0008)
Wind	9h	1.0517 (0.0060)	0.8933 (0.0038)	1.0053 (0.0114)	1.0190 (0.0041)	1.1419 (0.0049)	0.9157 (0.0020)
T2m	3h	0.3878 (0.0013)	0.3707 (0.0005)	0.3814 (0.0006)	0.3832 (0.0011)	0.4229 (0.0051)	0.3803 (0.0014)
T2m	6h	0.5630 (0.0026)	0.5284 (0.0022)	0.5469 (0.0028)	0.5516 (0.0040)	0.6255 (0.0083)	0.5392 (0.0019)
T2m	9h	0.4118 (0.0014)	0.3929 (0.0004)	0.4035 (0.0007)	0.4062 (0.0013)	0.4480 (0.0048)	0.4031 (0.0014)
Cloud	3h	0.08136 (0.00008)	0.07468 (0.00019)	0.07822 (0.00065)	0.07938 (0.00034)	0.08681 (0.00084)	0.07750 (0.00021)
Cloud	6h	0.12465 (0.00013)	0.11449 (0.00022)	0.11995 (0.00105)	0.12163 (0.00049)	0.13173 (0.00239)	0.11793 (0.00044)
Cloud	9h	0.08304 (0.00007)	0.07635 (0.00021)	0.07989 (0.00067)	0.08139 (0.00035)	0.08863 (0.00082)	0.07922 (0.00021)
Radiation	3h	19.136 (0.086)	17.865 (0.065)	18.585 (0.039)	18.739 (0.113)	20.262 (0.137)	18.523 (0.118)
Radiation	6h	17.151 (0.043)	15.960 (0.034)	16.637 (0.030)	16.850 (0.063)	18.081 (0.207)	16.483 (0.061)
Radiation	9h	25.549 (0.361)	23.077 (0.430)	24.878 (0.216)	25.062 (0.538)	26.030 (0.502)	23.361 (0.066)

Pairs include a missing endpoint; radiation uses valid daylight intervals. Signed increments retain physical units; eligible pairs vary with lag.

Table 18: **Same-product CERRA: temporal-increment RMSE.** Physical-unit means; sample SD across three training seeds appears below each mean.

		CFM				AFM	
Variable	Lag	Base	Full	Static	Reverse	Base	Full
Wind	3h	1.9625 (0.0156)	1.6891 (0.0086)	1.8831 (0.0252)	1.9011 (0.0213)	2.1386 (0.0114)	1.7323 (0.0055)
Wind	6h	2.3188 (0.0288)	1.7567 (0.0141)	2.1104 (0.0387)	2.1565 (0.0202)	2.5353 (0.0246)	1.8677 (0.0006)
Wind	9h	1.9677 (0.0156)	1.6950 (0.0086)	1.8885 (0.0252)	1.9143 (0.0194)	2.1436 (0.0114)	1.7381 (0.0056)
T2m	3h	0.7334 (0.0026)	0.7083 (0.0011)	0.7237 (0.0009)	0.7270 (0.0027)	0.7815 (0.0064)	0.7190 (0.0023)
T2m	6h	1.0483 (0.0042)	0.9970 (0.0038)	1.0251 (0.0037)	1.0328 (0.0036)	1.1340 (0.0096)	1.0090 (0.0026)
T2m	9h	0.7846 (0.0027)	0.7554 (0.0011)	0.7702 (0.0018)	0.7752 (0.0024)	0.8344 (0.0058)	0.7685 (0.0022)
Cloud	3h	0.18217 (0.00025)	0.17020 (0.00050)	0.17769 (0.00138)	0.17898 (0.00071)	0.19105 (0.00184)	0.17477 (0.00070)
Cloud	6h	0.21633 (0.00052)	0.20624 (0.00040)	0.21206 (0.00207)	0.21418 (0.00082)	0.22089 (0.00468)	0.18645 (0.00107)
Cloud	9h	0.18305 (0.00024)	0.17121 (0.00050)	0.17864 (0.00139)	0.18039 (0.00054)	0.19201 (0.00183)	0.17581 (0.00070)
Radiation	3h	42.839 (0.192)	40.439 (0.115)	42.110 (0.102)	42.213 (0.228)	43.822 (0.246)	41.103 (0.219)
Radiation	6h	36.338 (0.162)	34.488 (0.068)	35.793 (0.067)	35.912 (0.185)	37.239 (0.274)	34.946 (0.155)
Radiation	9h	46.941 (0.616)	43.965 (0.659)	46.852 (0.383)	46.912 (0.689)	46.904 (0.540)	43.237 (0.496)

Pairs include a missing endpoint; radiation uses valid daylight intervals. Signed increments retain physical units; eligible pairs vary with lag.

Table 19: **ERA5 → CERRA: temporal-increment CRPS.** Physical-unit means; sample SD across three training seeds appears below each mean.

		CFM				AFM	
Variable	Lag	Base	Full	Static	Reverse	Base	Full
Wind	3h	1.3024 (0.0034)	1.0955 (0.0008)	1.2336 (0.0029)	1.2597 (0.0025)	1.3160 (0.0022)	1.2679 (0.0045)
Wind	6h	1.8466 (0.0058)	1.3867 (0.0054)	1.7170 (0.0049)	1.7645 (0.0053)	1.8524 (0.0075)	1.7730 (0.0101)
Wind	9h	1.4491 (0.0029)	1.2214 (0.0018)	1.3654 (0.0042)	1.4228 (0.0025)	1.4751 (0.0030)	1.4068 (0.0052)
T2m	3h	0.5141 (0.0015)	0.4943 (0.0008)	0.4996 (0.0003)	0.5089 (0.0019)	0.5442 (0.0057)	0.5086 (0.0042)
T2m	6h	0.6956 (0.0012)	0.6516 (0.0009)	0.6672 (0.0017)	0.6858 (0.0025)	0.7694 (0.0104)	0.6809 (0.0106)
T2m	9h	0.7283 (0.0024)	0.6912 (0.0016)	0.6977 (0.0008)	0.7232 (0.0027)	0.7834 (0.0064)	0.7139 (0.0083)
Cloud	3h	0.12742 (0.00022)	0.12306 (0.00034)	0.12512 (0.00020)	0.12654 (0.00028)	0.13001 (0.00046)	0.12527 (0.00035)
Cloud	6h	0.15897 (0.00037)	0.14846 (0.00042)	0.15474 (0.00064)	0.15618 (0.00045)	0.16251 (0.00056)	0.15226 (0.00012)
Cloud	9h	0.16342 (0.00039)	0.15231 (0.00062)	0.15614 (0.00050)	0.16156 (0.00058)	0.16556 (0.00089)	0.15651 (0.00050)
Radiation	3h	33.594 (0.107)	31.164 (0.184)	31.871 (0.129)	33.016 (0.166)	33.622 (0.214)	31.557 (0.024)
Radiation	6h	42.938 (0.208)	39.185 (0.297)	40.053 (0.309)	42.069 (0.330)	41.520 (0.115)	39.193 (0.133)
Radiation	9h	49.951 (0.809)	47.955 (1.289)	49.891 (1.213)	49.931 (0.812)	53.429 (0.634)	49.031 (0.301)

Pairs include a missing endpoint; radiation uses valid daylight intervals. Signed increments retain physical units; eligible pairs vary with lag.

Table 20: **ERA5 → CERRA: temporal-increment RMSE.** Physical-unit means; sample SD across three training seeds appears below each mean.

		CFM				AFM	
Variable	Lag	Base	Full	Static	Reverse	Base	Full
Wind	3h	2.4293 (0.0037)	2.0330 (0.0039)	2.3186 (0.0098)	2.3469 (0.0082)	2.4630 (0.0042)	2.3895 (0.0090)
Wind	6h	3.3961 (0.0059)	2.5525 (0.0143)	3.2076 (0.0154)	3.2939 (0.0125)	3.4243 (0.0136)	3.3267 (0.0197)
Wind	9h	2.6220 (0.0062)	2.2187 (0.0058)	2.5049 (0.0128)	2.5648 (0.0113)	2.6794 (0.0036)	2.5873 (0.0113)
T2m	3h	0.9635 (0.0025)	0.9352 (0.0017)	0.9439 (0.0012)	0.9589 (0.0024)	1.0087 (0.0100)	0.9523 (0.0065)
T2m	6h	1.2946 (0.0015)	1.2307 (0.0017)	1.2543 (0.0028)	1.2782 (0.0018)	1.4113 (0.0204)	1.2762 (0.0196)
T2m	9h	1.3320 (0.0029)	1.2856 (0.0012)	1.2940 (0.0019)	1.3226 (0.0022)	1.4303 (0.0095)	1.3229 (0.0123)
Cloud	3h	0.26175 (0.00035)	0.25479 (0.00035)	0.25899 (0.00062)	0.26010 (0.00061)	0.26574 (0.00052)	0.25709 (0.00088)
Cloud	6h	0.31305 (0.00024)	0.29652 (0.00084)	0.30836 (0.00163)	0.30957 (0.00071)	0.31924 (0.00091)	0.30153 (0.00033)
Cloud	9h	0.31600 (0.00064)	0.29964 (0.00088)	0.30643 (0.00104)	0.31294 (0.00086)	0.31924 (0.00062)	0.30563 (0.00047)
Radiation	3h	66.310 (0.039)	63.013 (0.205)	64.390 (0.301)	65.704 (0.223)	65.343 (0.232)	62.894 (0.157)
Radiation	6h	84.685 (0.214)	80.360 (0.465)	81.833 (0.438)	83.733 (0.392)	82.139 (0.342)	79.289 (0.393)
Radiation	9h	89.623 (0.882)	88.361 (1.995)	89.585 (1.950)	89.611 (2.590)	96.194 (3.183)	88.525 (2.924)

Pairs include a missing endpoint; radiation uses valid daylight intervals. Signed increments retain physical units; eligible pairs vary with lag.

C.5 PAIRED MONTH-BLOCK INTERVALS AND TEMPORAL COVERAGE

Table 21 reports paired month-block intervals conditional on the three fitted models; seed SD captures training variation. Scores across lags, variables and metrics share cases and are correlated.

Six-hour instantaneous-variable scores cover 144 cases. Same-product CERRA radiation covers 126 cases/12 months at 3 h, 58/10 at 6 h and 7/4 at 9 h. The nine-hour Full–Base CRPS reduction is 9.673% ($[6.030, 14.753]\%$), with the narrowest weather coverage.

Validation-selected epochs (seeds 42/43/44) are CFM Base: 60/60/60; Full and Static: 35/35/40; Reverse: 35/45/45; AFM Base: 40/45/45; Full: 60/55/55. All use the same selection criterion.

Table 21: **Same-product CERRA: six-hour CRPS paired intervals.** Relative CRPS reductions and 95% CIs from 2,000 paired month-block bootstrap resamples.

Host	Comparator	Variable	Reduction	95% CI
CFM	Base	Cloud	+8.152%	[+6.289%, +9.321%]
CFM	Base	Radiation	+6.947%	[+3.190%, +10.438%]
CFM	Base	T2m	+6.156%	[+3.636%, +8.719%]
CFM	Base	Wind	+19.058%	[+16.629%, +21.240%]
AFM	Base	Cloud	+10.479%	[+8.981%, +12.459%]
AFM	Base	Radiation	+8.838%	[+4.521%, +12.521%]
AFM	Base	T2m	+13.791%	[+10.559%, +17.666%]
AFM	Base	Wind	+24.206%	[+21.996%, +27.983%]
CFM	Static	Cloud	+4.558%	[+3.511%, +6.123%]
CFM	Static	Radiation	+4.072%	[+0.404%, +7.383%]
CFM	Static	T2m	+3.385%	[+1.816%, +5.181%]
CFM	Static	Wind	+14.571%	[+13.148%, +16.379%]
CFM	Reverse	Cloud	+5.877%	[+4.874%, +7.293%]
CFM	Reverse	Radiation	+5.286%	[+1.728%, +8.185%]
CFM	Reverse	T2m	+4.211%	[+2.569%, +6.315%]
CFM	Reverse	Wind	+15.572%	[+14.422%, +18.155%]

Resampling units are months; the three fitted models remain fixed.

C.6 ACROSS-SCALE RECONSTRUCTION

Table 22 reports missing-time block-mean CRPS and single-time spatial-detail diagnostics over the full test sets. Full improves block-mean, detail and wind spatial-increment CRPS over Base in every displayed comparison for both hosts.

Table 22: **Across-scale reconstruction on the full test sets.** CRPS of final generated samples; lower is better.

Diagnostic	CFM Base	CFM Full	AFM Base	AFM Full
East Asian ERA5				
Missing-time block-mean CRPS	0.097500	0.068300	0.105900	0.080400
Wind detail CRPS	0.469584	0.422468	0.519346	0.451378
Wind spatial x-increment CRPS	0.242416	0.221987	0.290258	0.237797
Cloud detail CRPS	0.054183	0.052524	0.056973	0.054159
Radiation detail CRPS	7.458361	7.214175	7.924061	7.543673
ERA5 → CERRA				
Missing-time block-mean CRPS	0.196200	0.134700	0.208400	0.156900
Wind detail CRPS	1.594700	1.356500	1.624800	1.384900
Wind spatial x-increment CRPS	0.324700	0.294000	0.351900	0.313600
Cloud detail CRPS	0.064800	0.051400	0.068100	0.054200
Radiation detail CRPS	13.874000	9.316000	14.592000	10.487000

Seed SD is unavailable for these diagnostics.

D QUALITATIVE RECONSTRUCTION

State panels show raw 12-member ensemble means from validation-selected seed-42 checkpoints at NFE 30. The local CRPS-difference map scores the full ensembles.

D.1 ERROR AT UNRESOLVED COARSE AND FINE SCALES

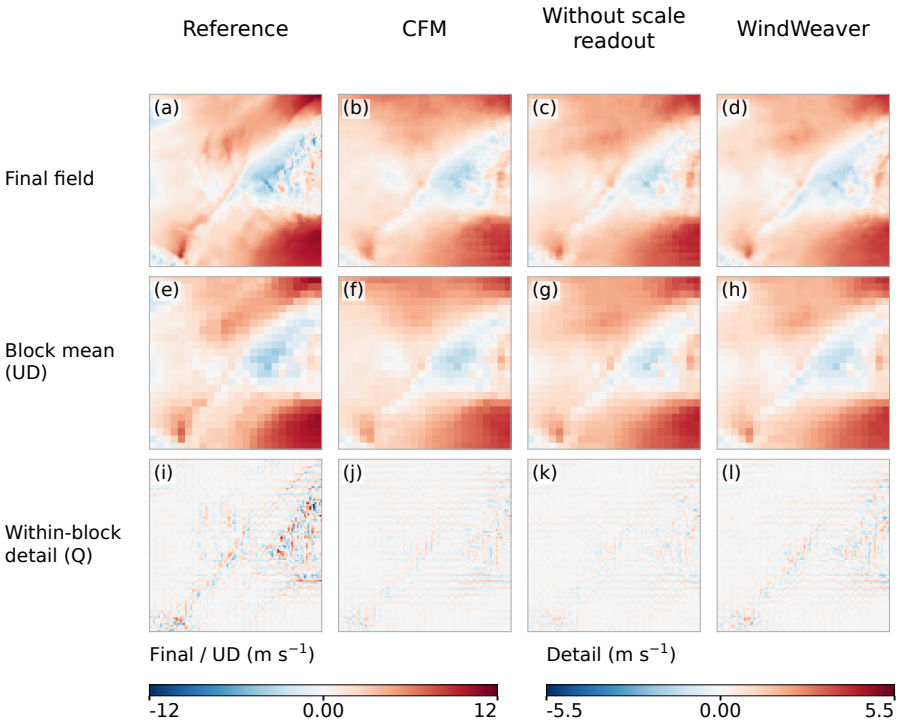

Figure 5: **Final-field scale decomposition in January.** The missing 15 UTC u_{850} state in the same-product CERRA case of Figure 3. Each final ensemble-mean field is decomposed with the task's area-weighted 4×4 aggregation: $x = UDx + Qx$. WindWeaver reduces both coarse-state and detail errors in this case.

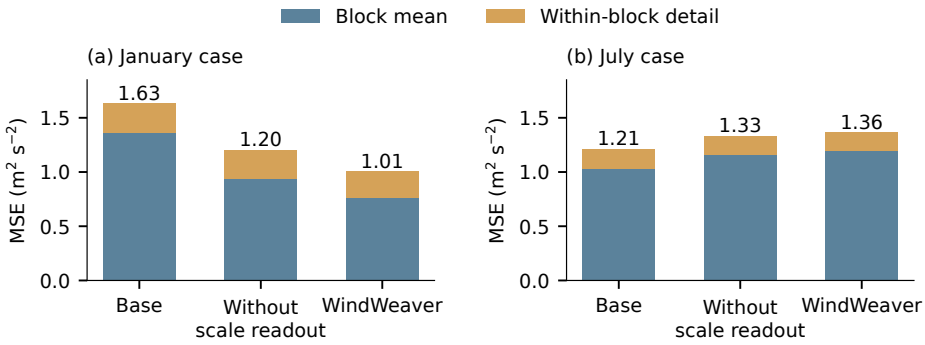

Figure 6: **The scale of improvement varies by case.** Area-weighted u_{850} ensemble-mean MSE at the missing 3 h state. Orthogonality gives total MSE as the sum of block-mean and detail MSE. January's reduction is predominantly coarse-scale; July retains larger coarse-state error. Each bar partitions final-field error by spatial scale for one case and seed.

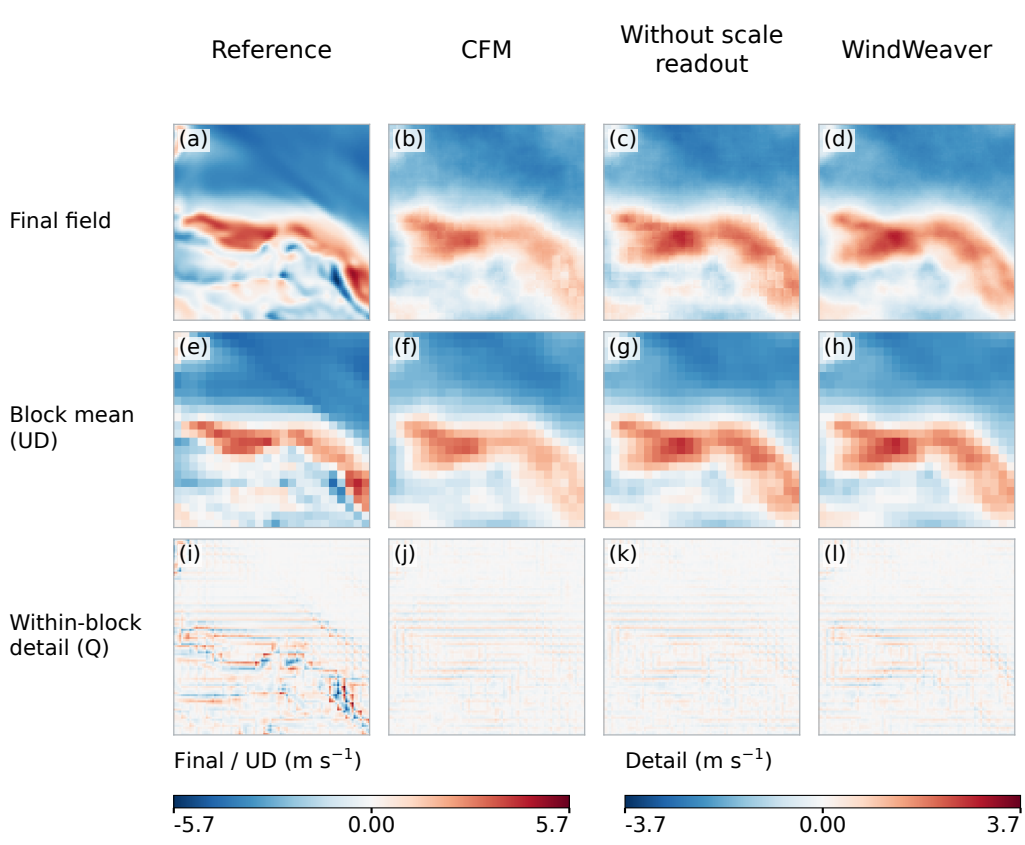

Figure 7: **Final-field scale decomposition in July.** Same-product CERRA, 16 July 2022 at 15 UTC, in the patch spanning approximately 30.7–35.5°N, 10.3–16.1°E. Decomposition follows Figure 5; color scales are shared across methods within this case. Relative to Base, WindWeaver has slightly lower detail RMSE but larger block-mean and final-field RMSE. Narrow reference structures remain unresolved.

The two examples illustrate distinct reconstruction outcomes. At 3 h in January, Base, the configuration without scale readouts, and WindWeaver have final-field RMSE of 1.277, 1.097 and 1.003 m/s. In July these are 1.100, 1.155 and 1.168 m/s. The examples locate errors within ensemble means; Table 22 evaluates the full ensembles across the test set.

D.2 Matched reconstruction and error maps

Each following pair shows reference and reconstructed states above, and the reference with four absolute-error fields below. State scales are shared across methods, and errors use a separate shared scale. Temperature states use °C and errors use K; cloud states use % and errors use percentage points. Radiation is mean flux over the preceding 1 h for ERA5 and 3 h for CERRA. European fields are displayed on the native 5.5 km grid. Purple marks out-of-range cloud values and grey marks unavailable values. Maps retain raw ensemble means; tabulated probability scores use the full ensembles.

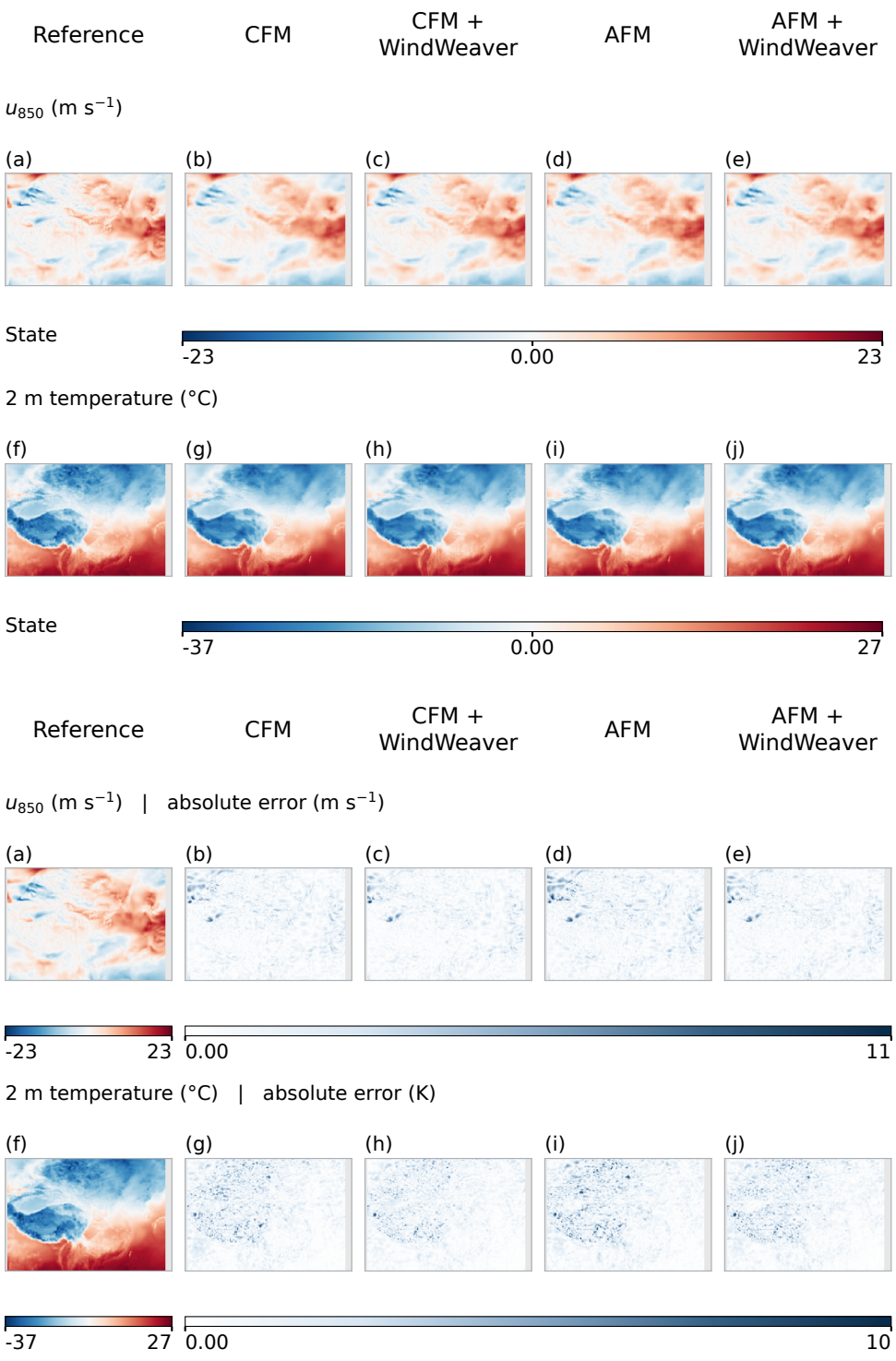

Figure 8: **East Asian ERA5: January wind and temperature.** 16 January 2022 at 03 UTC, from a 00 UTC window. Spatial resolution changes from $1.5°$ to $0.25°$ and temporal spacing from 6 h to 1 h.

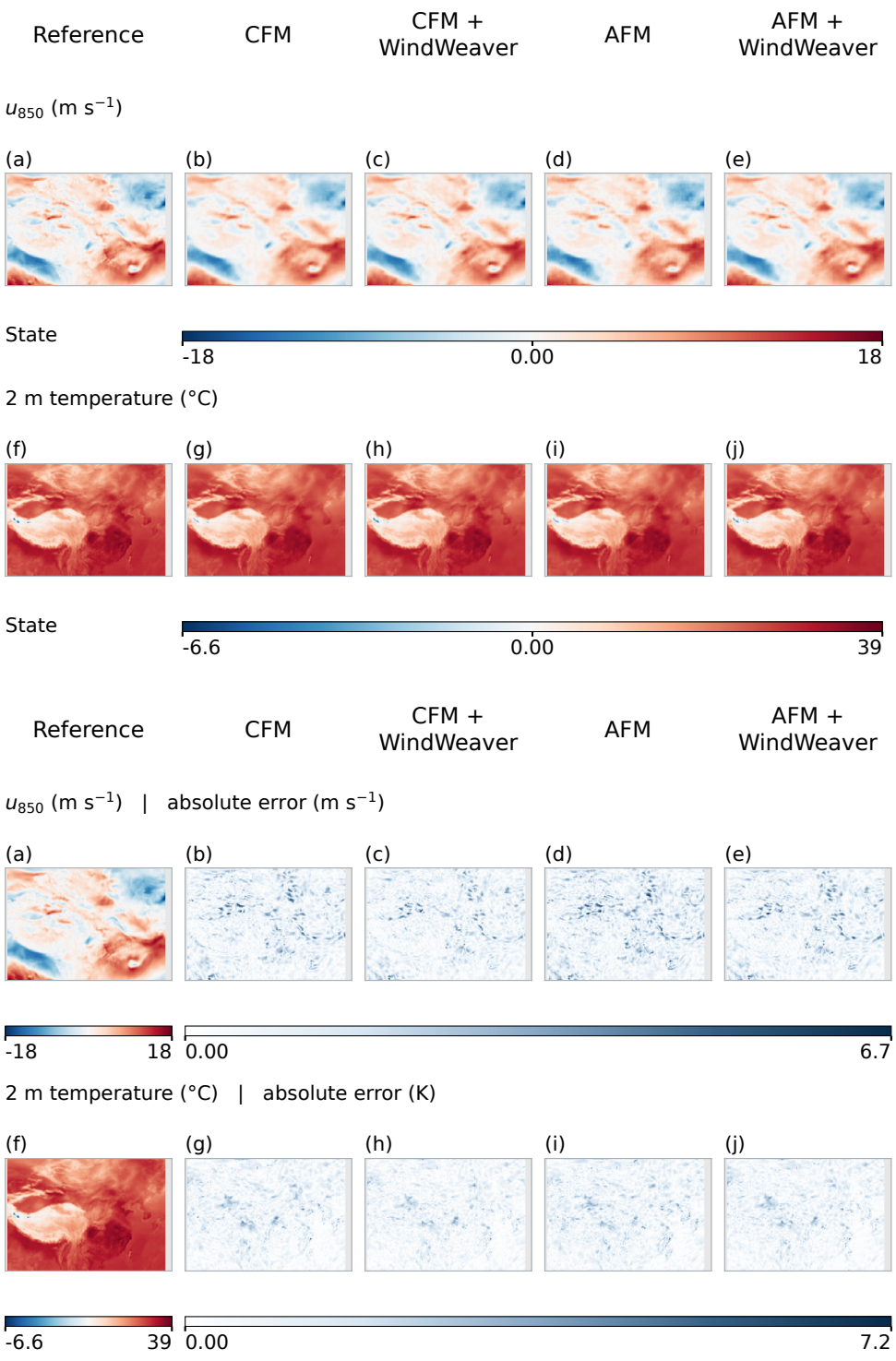

Figure 9: **East Asian ERA5: July wind and temperature.** 16 July 2022 at 03 UTC, with the same protocol as Figure 8.

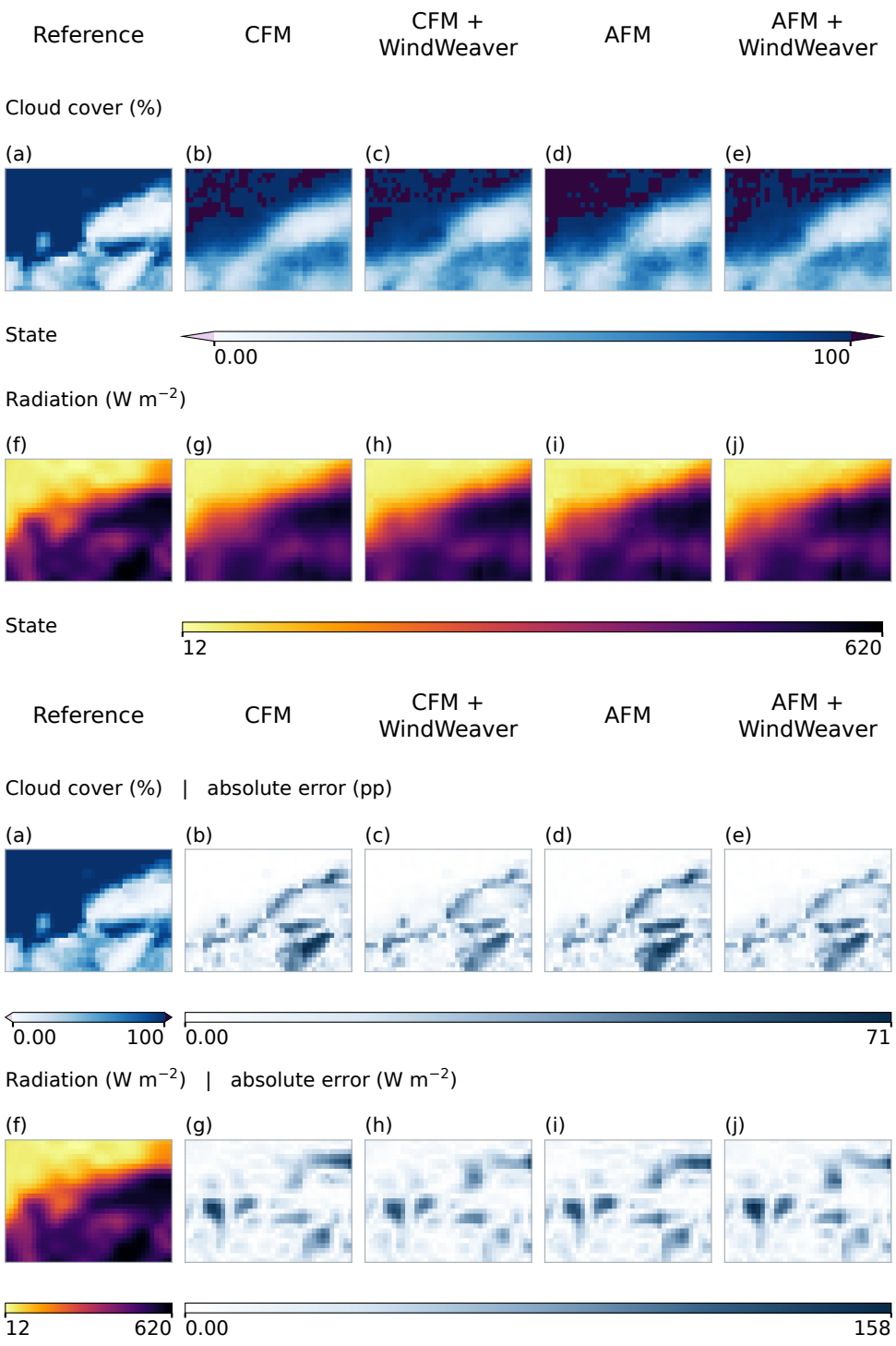

Figure 10: **Cloud and radiation in Guangdong and surroundings.** The East Asian ERA5 model January window, cropped to the fixed rectangle 109–118°E, 20–26°N. The crop uses the East Asian models and a fixed geographic extent.

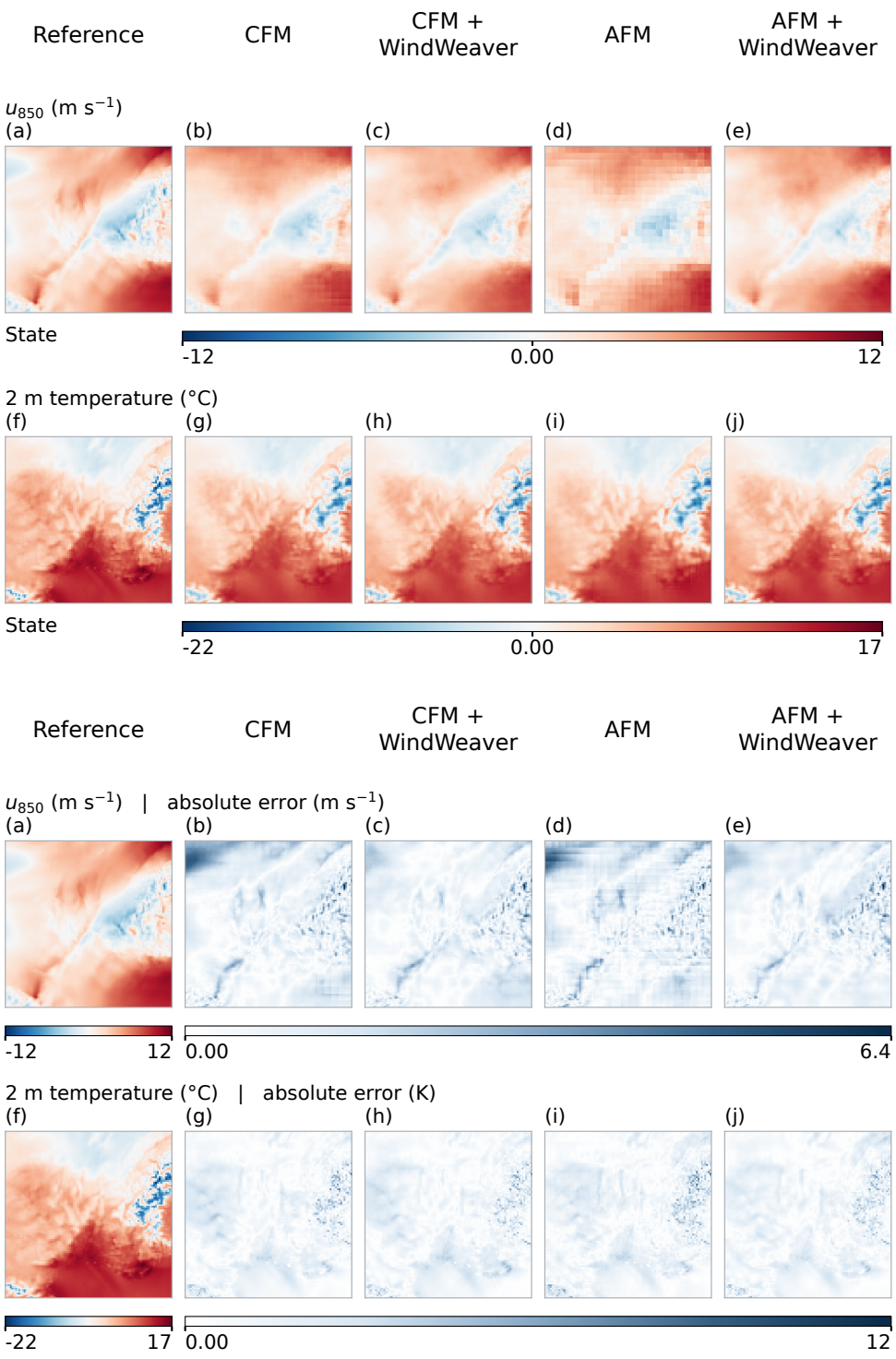

Figure 11: **Same-product CERRA: January wind and temperature.** 16 January 2022 at 15 UTC, three hours after the window start. Inputs are 22 km/6 h and targets 5.5 km/3 h. The patch matches Figure 3.

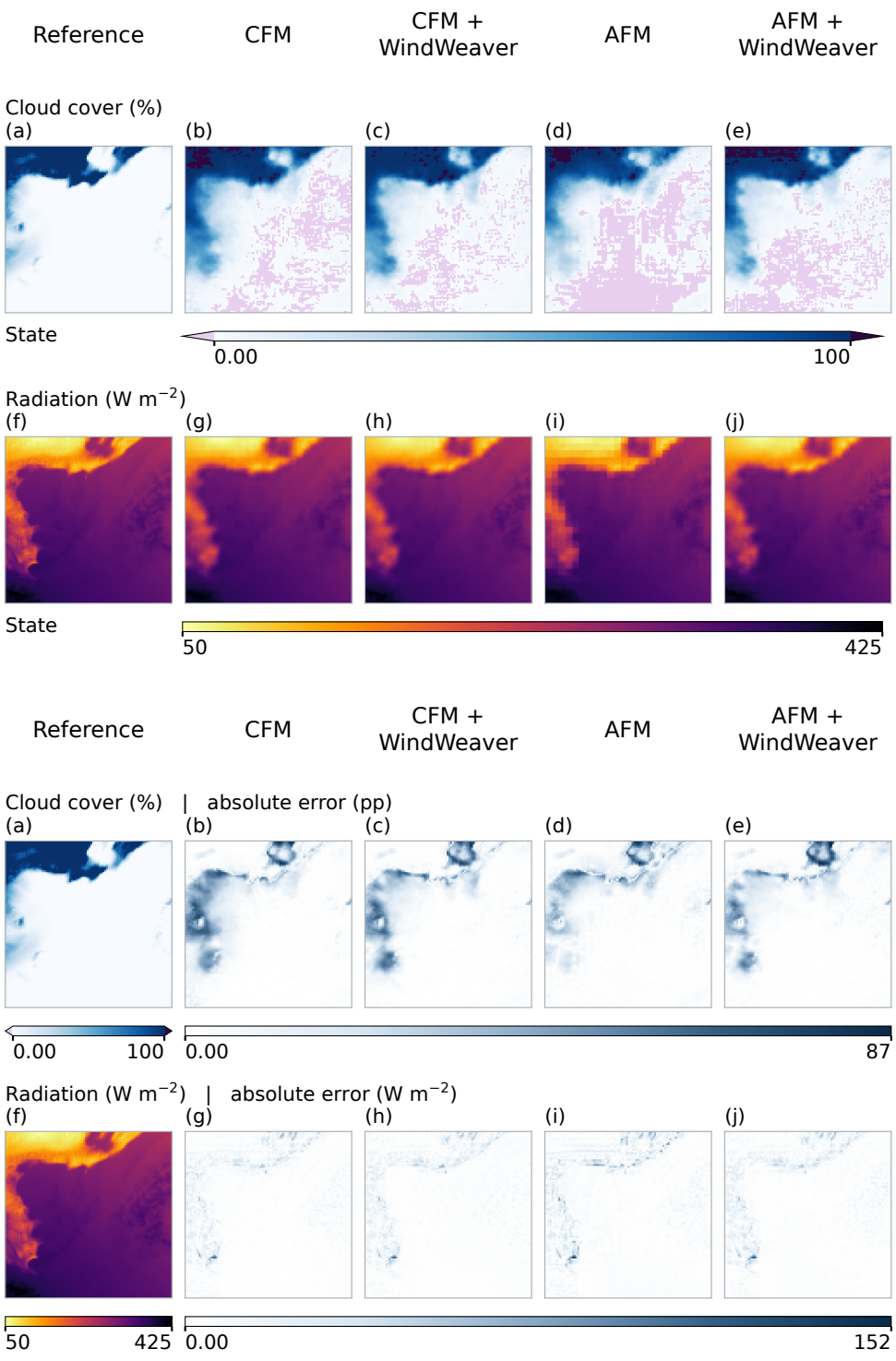

Figure 12: **Same-product CERRA: January cloud and radiation.** The same missing 15 UTC state as Figure 11. Radiation covers 12–15 UTC. Local gains differ by variable and host.

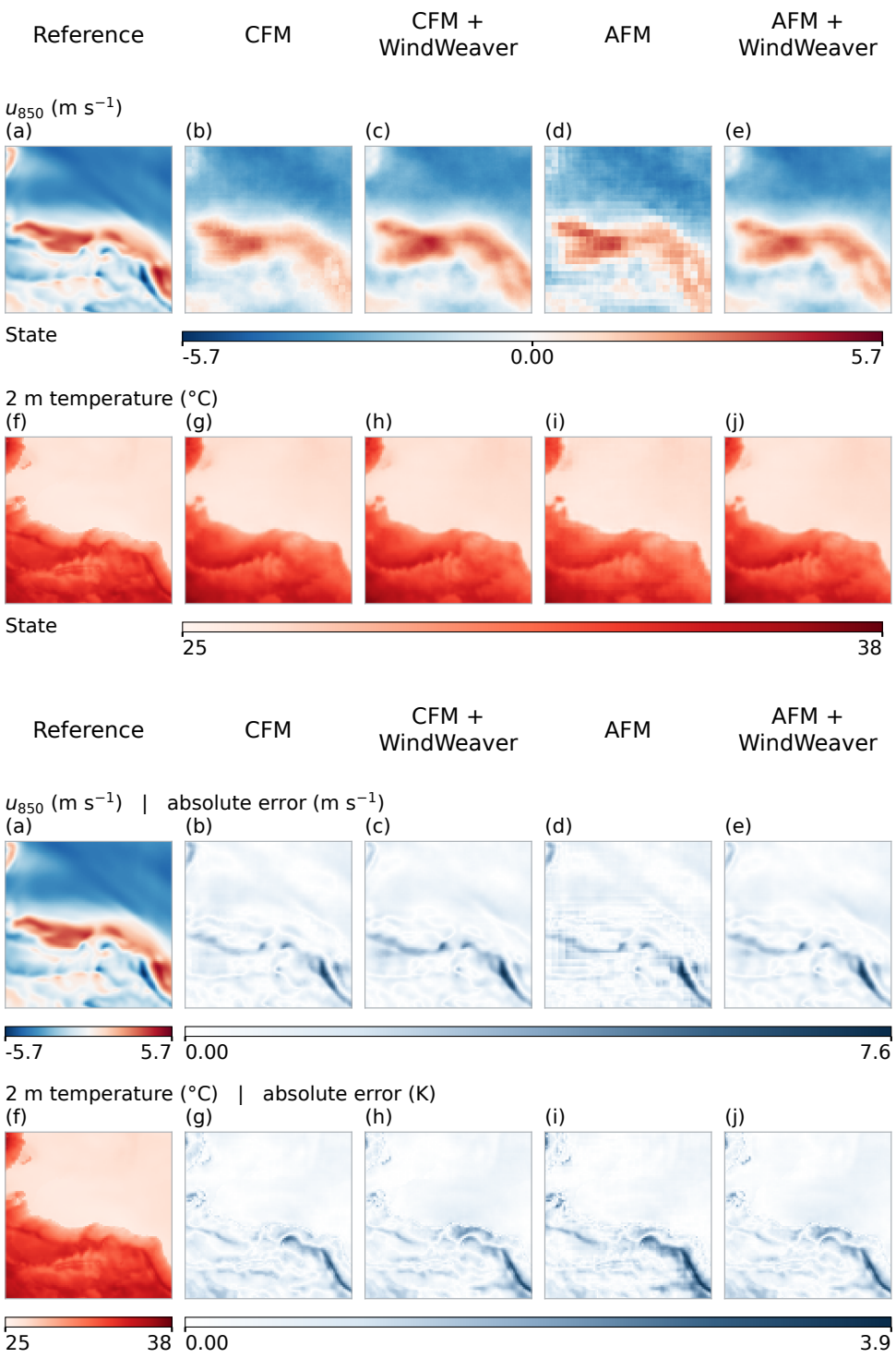

Figure 13: **Same-product CERRA: July wind and temperature.** 16 July 2022 at 15 UTC. Figure 7 decomposes this case's CFM wind-state error, which is larger with WindWeaver than with Base.

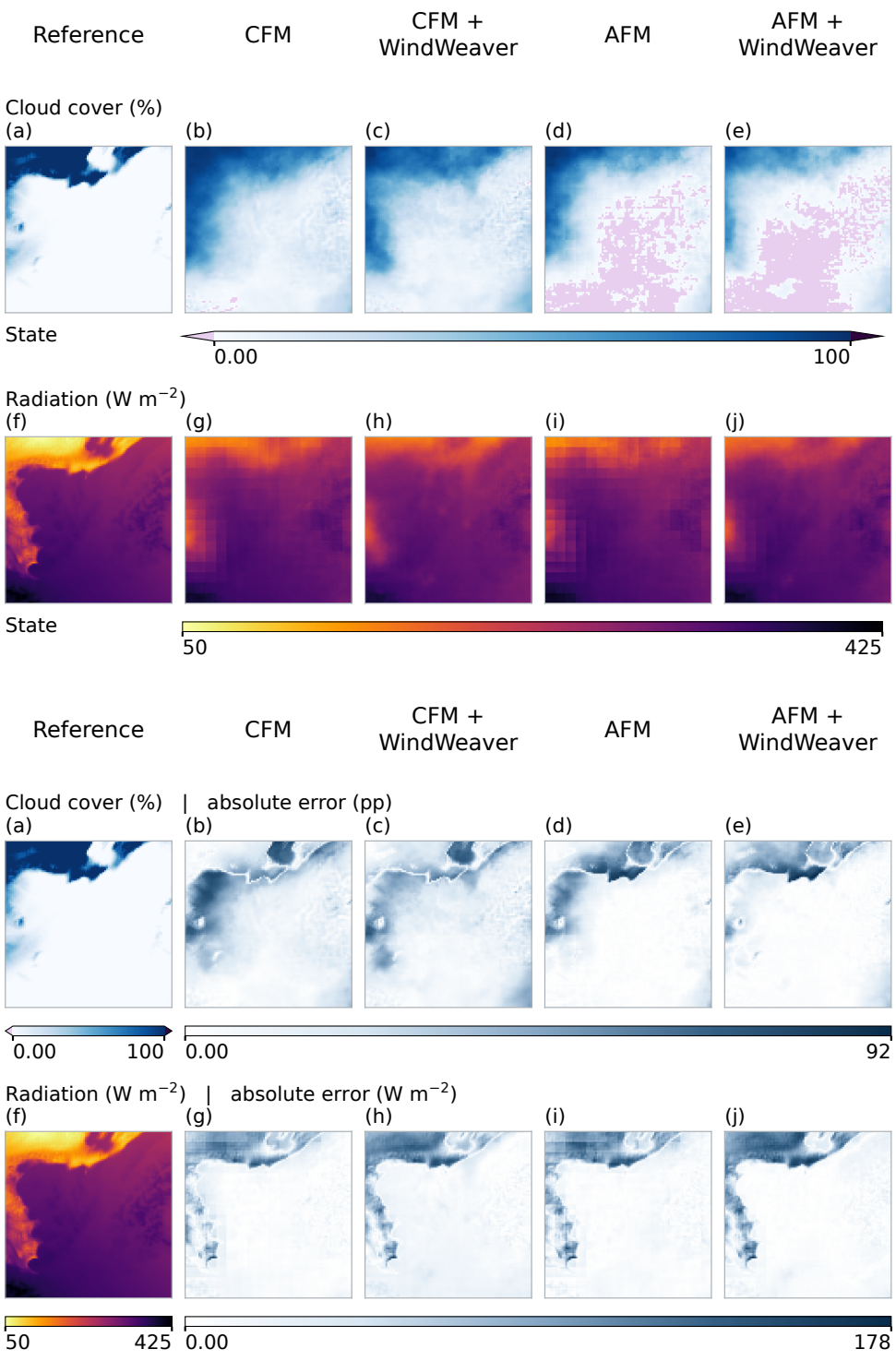

Figure 14: **Paired ERA5–CERRA: January cloud and radiation.** 16 January 2022 at 15 UTC. Native ERA5 conditions are interpolated to 33 km centers at 6 h spacing; CERRA targets are 5.5 km/3 h. This paired-product task allows target-product mean corrections.

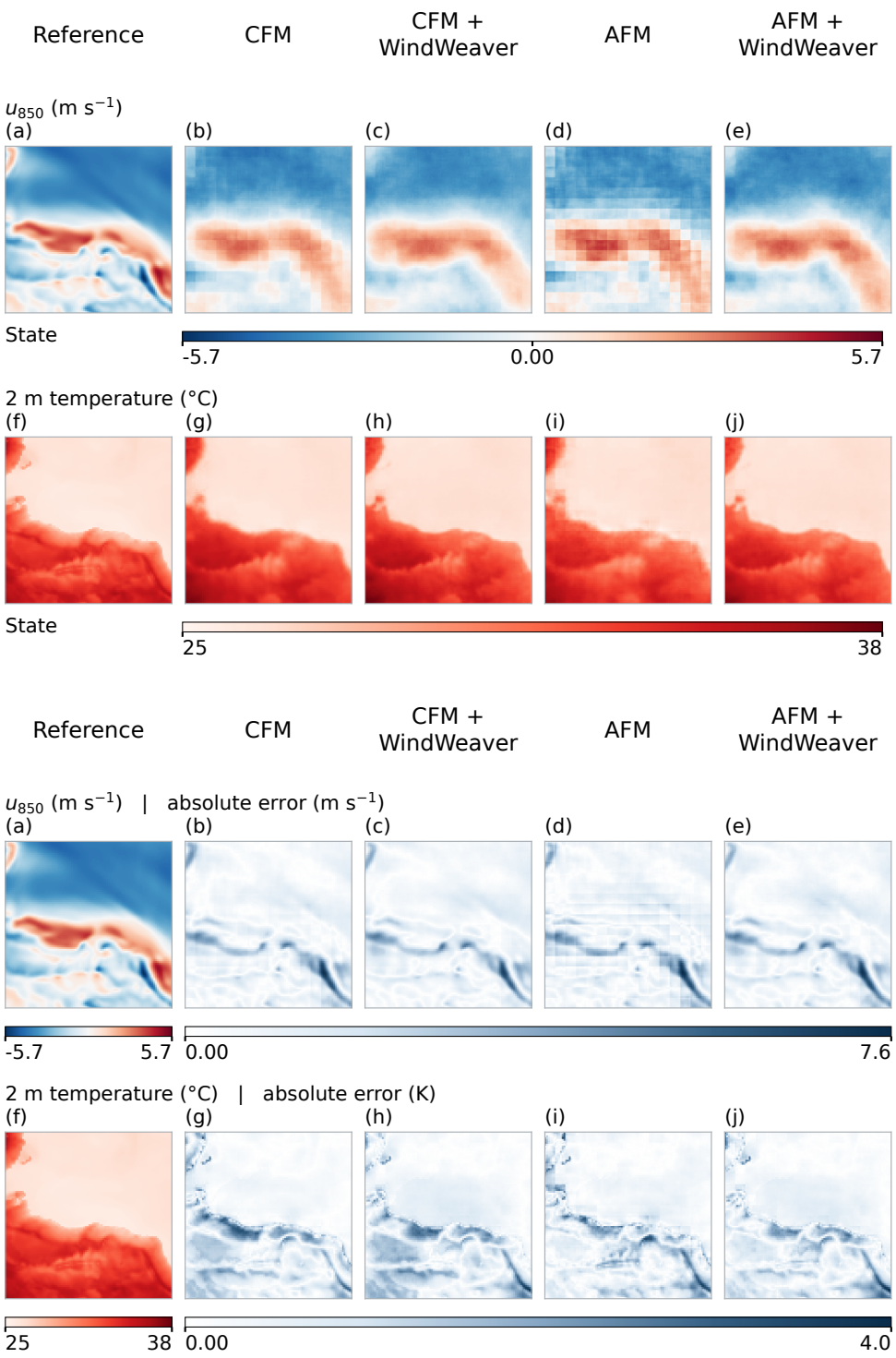

Figure 15: **Paired ERA5–CERRA: July wind and temperature.** 16 July 2022 at 15 UTC, in the patch spanning approximately 30.7–35.5°N, 10.3–16.1°E. Block-like artifacts are reduced in some reconstructed fields, while fine reference structures remain unresolved.

D.3 Spatial variation of probabilistic gains

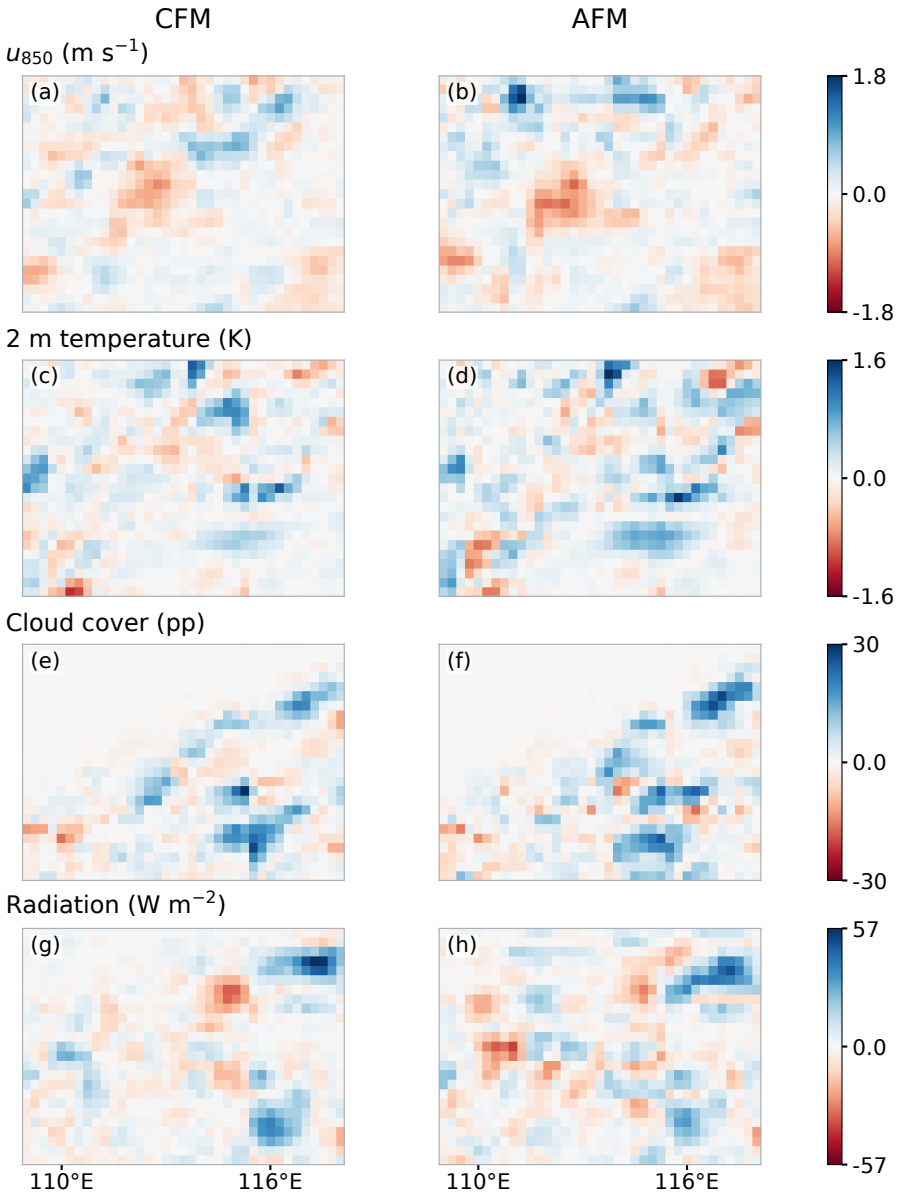

Figure 16: **Local gains in an East Asian ERA5 winter case.** Pointwise $\mathrm{CRPS}(\mathrm{Base}) - \mathrm{CRPS}(\mathrm{WindWeaver})$ on 16 January 2022 at 03 UTC, from the East Asian ERA5 models in the fixed rectangle 109–118°E, 20–26°N. Blue indicates improvement and red deterioration; each variable uses one symmetric scale for both hosts. Wind denotes u_{850}. The maps show one test time and seed. Cloud gains concentrate in parts of the region, while wind has mixed local changes.

E TRAINING TRAJECTORIES AND COMPUTATIONAL COST

We report the recorded validation trajectories and computational costs of CFM Base, Full, and No-exchange on East Asian ERA5 and ERA5→CERRA. These records cover 18 existing training runs: three seeds for each method and protocol.

E.1 VALIDATION TRAJECTORIES

Figure 17 shows all 216 recorded validation scores. Each run completes the same 60-epoch budget of 36,000 optimizer updates. Validation evaluates the EMA model every five epochs on the fixed validation split, using four ensemble members and 20 Euler steps. The original minimum-validation-CRPS checkpoint is epoch 60 in all 18 runs.

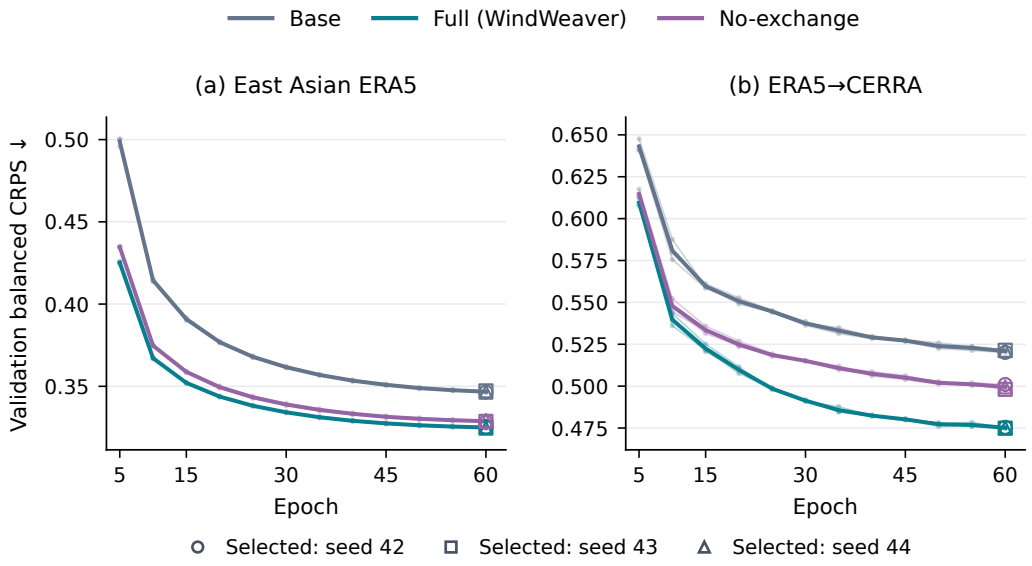

Figure 17: **Validation trajectories under the matched training budget.** Panels show East Asian ERA5 and ERA5→CERRA. Faint lines show seeds 42–44; solid lines are their arithmetic means at each recorded epoch. Open markers identify the original selected checkpoints. Validation uses four members and 20 Euler steps, whereas standard testing uses 12 members and 30 steps. Scores retain the four-group normalization and protocol-specific masks. Curves are unsmoothed, with separate vertical scales for the two protocols.

Full has lower validation balanced CRPS than Base and No-exchange at every recorded checkpoint for each seed in both protocols. Its advantage persists across the observed training trajectory, with all minima occurring at the final recorded epoch. Longer training would be needed to assess convergence beyond this budget. Validation and test scores use their respective windows and sampling settings.

E.2 RECORDED TRAINING AND INFERENCE COSTS

Table 23: **Recorded costs for CFM variants.** Training reports hours per complete 60-epoch run, with mean ± sample SD over three training seeds. Inference reports seconds per 12-member window, with mean ± measurement SD over 12 timed calls. Memory is the maximum peak PyTorch allocation across those calls. Training hardware history is unavailable; inference uses a shared A100 PCIe 40GB.

Protocol	CFM variant	Training (h)	Inference (s)	Allocated (GiB)
East Asian ERA5	Base	1.907 ± 0.013	3.946 ± 0.527	1.695
	Full	2.057 ± 0.010	4.123 ± 0.517	1.724
	No-exchange	1.942 ± 0.012	3.837 ± 0.022	1.712
ERA5→CERRA	Base	2.904 ± 0.037	5.455 ± 0.199	2.595
	Full	3.071 ± 0.028	5.865 ± 0.194	2.643
	No-exchange	2.934 ± 0.033	5.505 ± 0.190	2.624

Training. Training time in Table 23 is the recorded wall time through primary-test completion minus the separately timed primary test. It covers initialization, all 60 epochs, data handling, validation, checkpoint operations, final EMA reloading, and bookkeeping, ending before subsequent NFE evaluations. Each run uses one GPU; historical records omit its identity and external load.

Inference. We retain the original seed-42 EMA checkpoints and sampling path, with 12 members, 30 Euler steps per member, and four-member chunks at window batch size one. Measurements use one shared NVIDIA A100 PCIe 40GB GPU with PyTorch 2.4.0 and CUDA 12.4. Precision settings are FP32 states and weights, BF16 backbone autocast, and enabled TF32 matrix multiplication. Each method and protocol has 12 timed measurements: four fixed cases repeated over three blocks, with three complete warm-ups per block. Method order is interleaved across blocks.

CUDA synchronization brackets each timed call. Timing includes source generation, conditioning, motion geometry, flow integration, reconstruction, and output assembly. It excludes model loading, host-to-device transfer, disk I/O, warm-up, and output writing. Each member reconstructs a complete $9 \times 13 \times 96 \times 96$ window for East Asian ERA5 or $9 \times 5 \times 192 \times 192$ window for ERA5→CERRA (channels, weather times, height, width), before any central scoring crop.

Peak-memory statistics are reset before each measurement; reported PyTorch allocation covers the resident model, input/reference tensors, and generated outputs. Checkpoint SHA256 records match before and after measurement. External GPU workloads were present during timing, so the table reports absolute observed costs under shared load. Controlled speed comparisons would require matched hardware and load conditions.