

Supplementary Information: FuXi-Nowcast

Lei Chen^{1†}, Zijian Zhu^{1,5†}, Xiaoran Zhuang^{2,3†}, Tianyuan Qi¹,
Yuxuan Feng^{2,3}, Xiaohui Zhong^{1,4*}, Hao Li^{1,4,5*}

^{1*}Artificial Intelligence Innovation and Incubation Institute, Fudan
University, Shanghai, 200433, China.

²Jiangsu Meteorological Observatory, Nanjing, Jiangsu, 210008, China.

³Jiangsu Key Laboratory of Severe Storm Disaster Risk / Key
Laboratory of Transportation Meteorology of CMA, Nanjing, Jiangsu,
210008, China.

⁴FuXi Intelligent Computing Technology Co., Ltd., Shanghai, 200233,
China.

⁵Shanghai Innovation Institute, Shanghai, China.

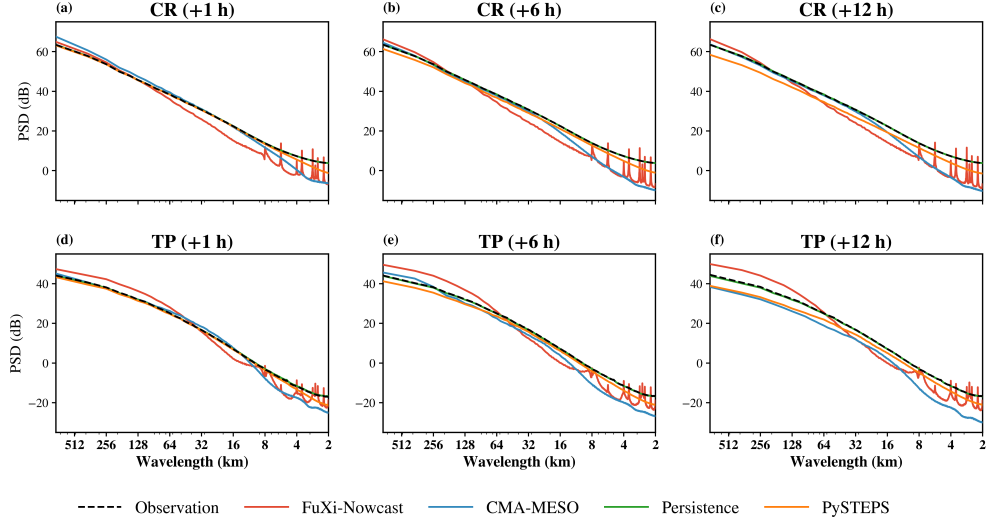
*Corresponding author(s). E-mail(s): x7zhong@gmail.com;
lihao_lh@fudan.edu.cn;

Contributing authors: cltpys@163.com; 24210240443@m.fudan.edu.cn;
zxrxz3212009@163.com; 25213050312@m.fudan.edu.cn;
fengyuxuan27@163.com;

[†]These authors contributed equally to this work.

Additional verification results

The power spectral density comparison in Supplementary Figure 1 shows that FuXi-Nowcast broadly follows the observed spectral shape at large spatial scales, although it overestimates the large-scale spectral power of TP. At intermediate-to-short wavelengths, particularly around 16–64 km, FuXi-Nowcast exhibits reduced spectral power, and this deficit generally becomes larger with increasing lead time.



Supplementary Figure 1: Power spectral density across forecast methods and lead times. The radially averaged power spectral density (PSD) spectra for composite reflectivity (CR; panels (a)–(c)) and total precipitation (TP; panels (d)–(f)) from gridded observations, FuXi-Nowcast, CMA-MESO, Persistence, and PySTEPS at forecast lead times of +1, +6, and +12 h (columns). Black dashed curves denote gridded observations; red, blue, green, and orange solid curves denote FuXi-Nowcast, CMA-MESO, Persistence, and PySTEPS, respectively. Wavelength is shown on the horizontal axis, and the vertical axis shows PSD values in decibels (dB).

The main text uses CSI as the primary verification metric. Here we report the corresponding BIAS, FAR, and POD for the same variables and thresholds (Supplementary Figures 2–4) to further characterize model behavior.

BIAS (Supplementary Figure 2) quantifies systematic overprediction or underprediction. For FuXi-Nowcast, BIAS is comparatively stable, although deviations from the ideal value of 1.0 depend on the variable, threshold, and lead time. The China Meteorological Administration Mesoscale Model (CMA-MESO) shows pronounced overforecasting for GS, especially at longer lead times, consistent with its low CSI in the main text. As CMA-MESO GS is diagnosed from the 10-meter wind speed using a fixed factor of 1.75, its positive bias is likely attributable to this diagnostic conversion. The negative biases for both TP and CR at longer lead times suggest that CMA-MESO underestimates the frequency or spatial extent of the corresponding convective events. PySTEPS exhibits increasingly low BIAS with lead time, reflecting the progressive weakening of purely extrapolated echoes.

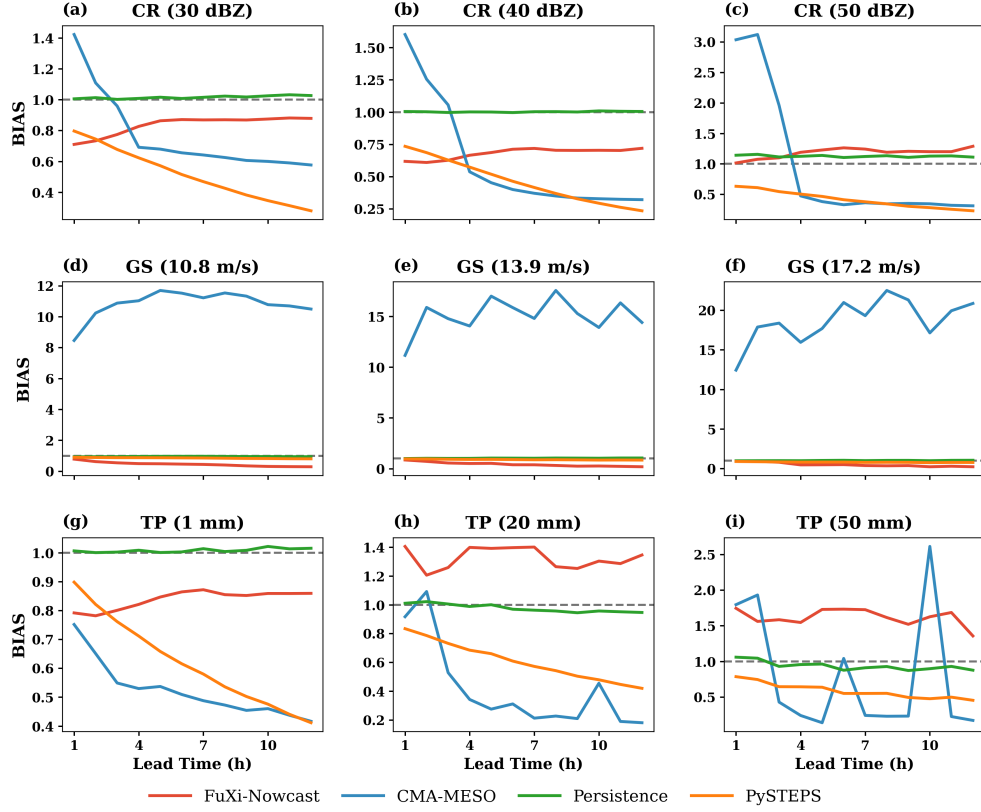
FAR (Supplementary Figure 3) shows that CMA-MESO generally has the lowest FAR for CR. This low FAR occurs together with a pronounced CR underforecasting bias and lower POD, indicating that it largely reflects too few forecast CR events rather than superior discrimination. For GS, CMA-MESO has the highest FAR, consistent

with its strong overforecasting bias, whereas FuXi-Nowcast has lower FAR than the other methods.

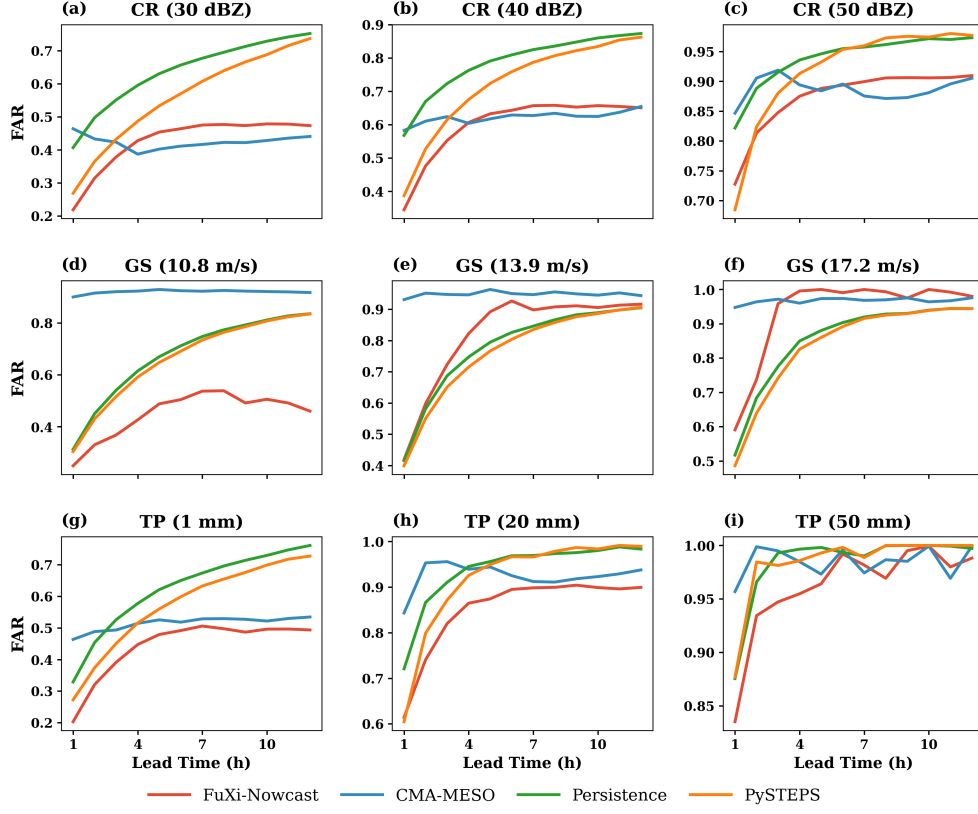
POD (Supplementary Figure 4) shows that the low CR FAR of CMA-MESO is accompanied by lower detection rates and an underforecasting bias. FuXi-Nowcast generally detects a larger fraction of observed CR events while maintaining a more balanced event frequency, which yields the higher CSI. For GS, the high event frequency in CMA-MESO is accompanied by a high FAR.

FSS (Supplementary Figures 5–8) quantifies spatial agreement while allowing limited displacement within a neighborhood. FSS generally increases with neighborhood size for all methods. For CR, FuXi-Nowcast generally achieves the highest FSS at the 30 and 40 dBZ thresholds and small-to-intermediate neighborhood scales, while CMA-MESO and PySTEPS become more competitive at larger scales and higher thresholds. For GS, FuXi-Nowcast shows the clearest advantage at 10.8 m s^{-1} , whereas all methods exhibit much lower FSS at higher thresholds and CMA-MESO becomes relatively competitive at longer lead times. For TP, FuXi-Nowcast generally maintains the highest FSS at the 1 and 20 mm thresholds, while PySTEPS often has higher short-lead FSS but decays rapidly with lead time. Overall, the FSS results are consistent with the CSI-based conclusions, while indicating that method rankings depend on both intensity threshold and neighborhood scale.

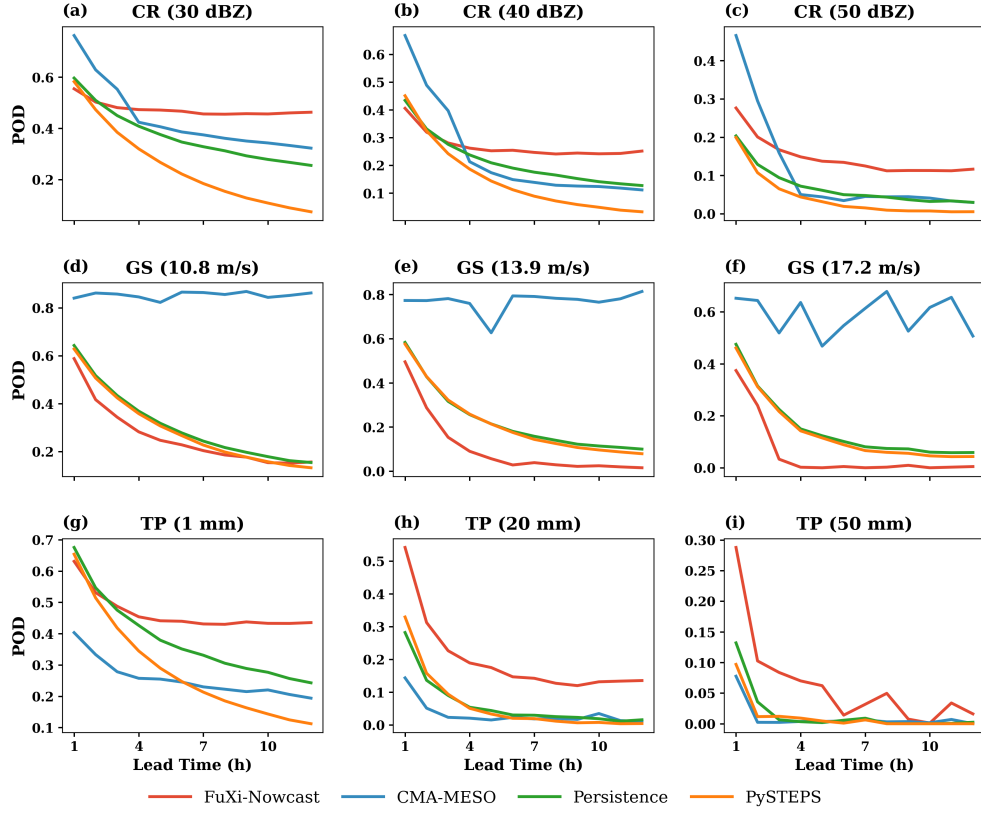
Overall, these metrics show that FuXi-Nowcast maintains a more balanced combination of event frequency, detection, and false alarms. For CMA-MESO, the low CR FAR occurs together with underforecasting and lower POD, whereas its high GS event frequency is accompanied by a high FAR. PySTEPS progressively loses skill as extrapolated echoes weaken. The FSS results further show that model rankings depend on both event intensity and neighborhood scale.



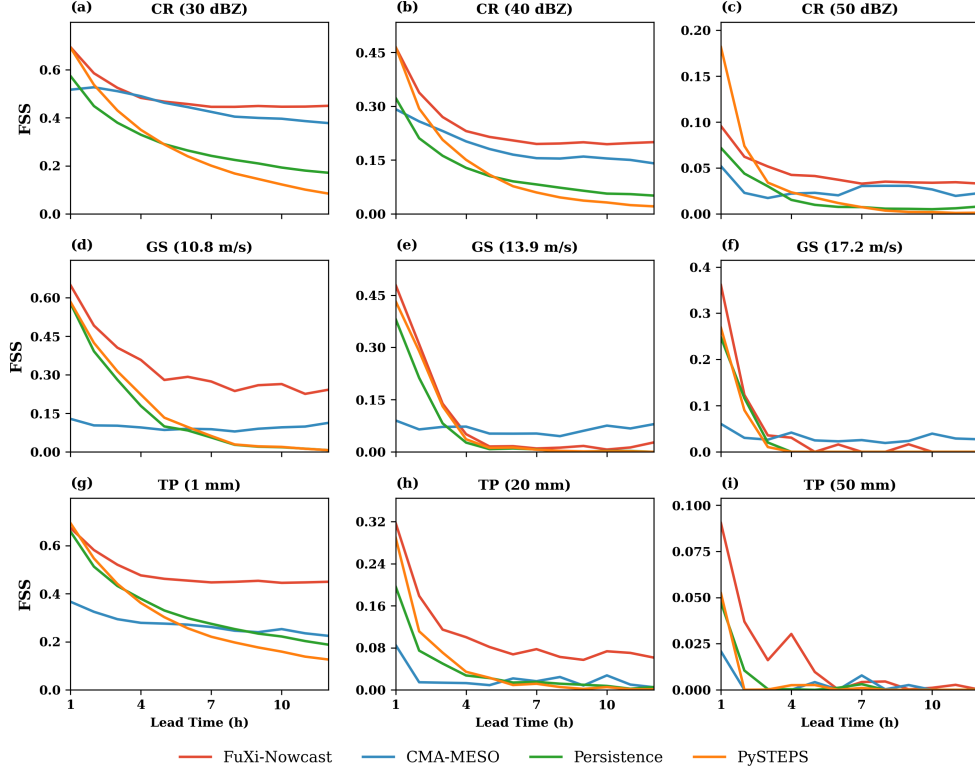
Supplementary Figure 2: Frequency bias across forecast methods, variables, and thresholds. Frequency bias (BIAS) comparison among FuXi-Nowcast (red), CMA-MESO (blue), Persistence (green), and PySTEPS (orange) for 12 h forecasts. Panels (a)–(c), (d)–(f), and (g)–(i) correspond to composite reflectivity (CR), wind gust (GS), and total precipitation (TP), respectively. Columns show three intensity thresholds for each variable. The grey dashed line indicates the perfect BIAS value of 1.0. Values above 1 indicate overforecasting; values below 1 indicate underforecasting.



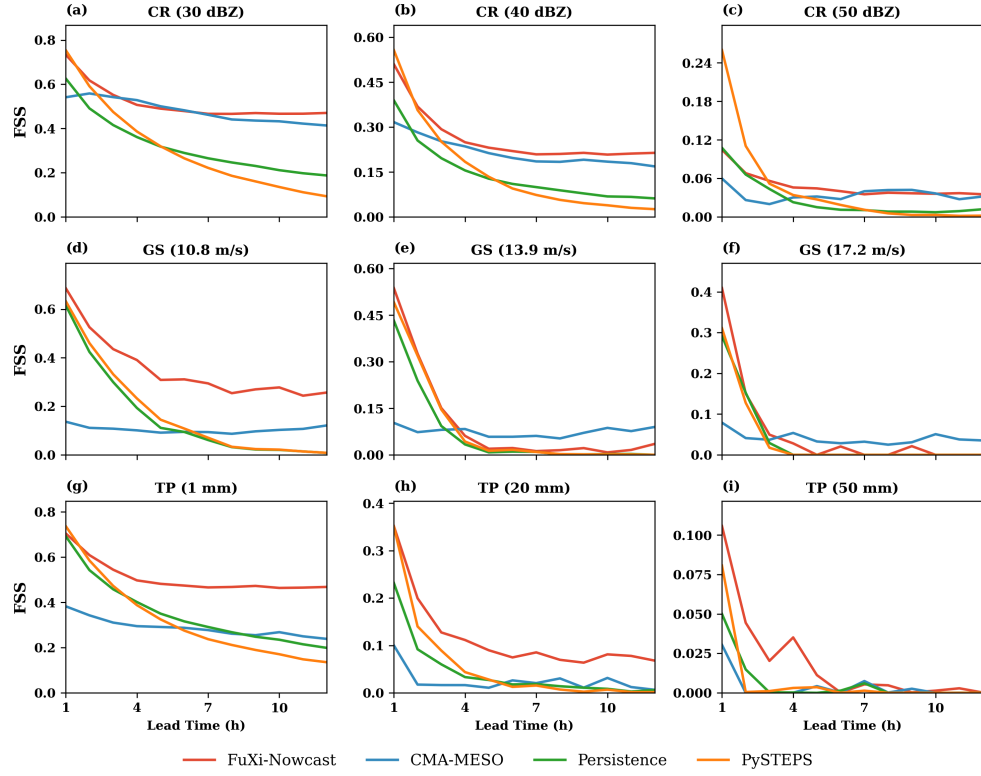
Supplementary Figure 3: False-alarm ratio across forecast methods, variables, and thresholds. The false-alarm ratio (FAR) is compared for FuXi-Nowcast (red), CMA-MESO (blue), Persistence (green), and PySTEPS (orange) for 12 h forecasts. Panels (a)–(c), (d)–(f), and (g)–(i) correspond to composite reflectivity (CR), wind gust (GS), and total precipitation (TP), respectively. Columns show CR thresholds of 30, 40, and 50 dBZ; GS thresholds of 10.8, 13.9, and 17.2 m s^{-1} ; and TP thresholds of 1, 20, and 50 mm. Lower FAR indicates fewer false alarms, and a value of zero indicates no false alarms.



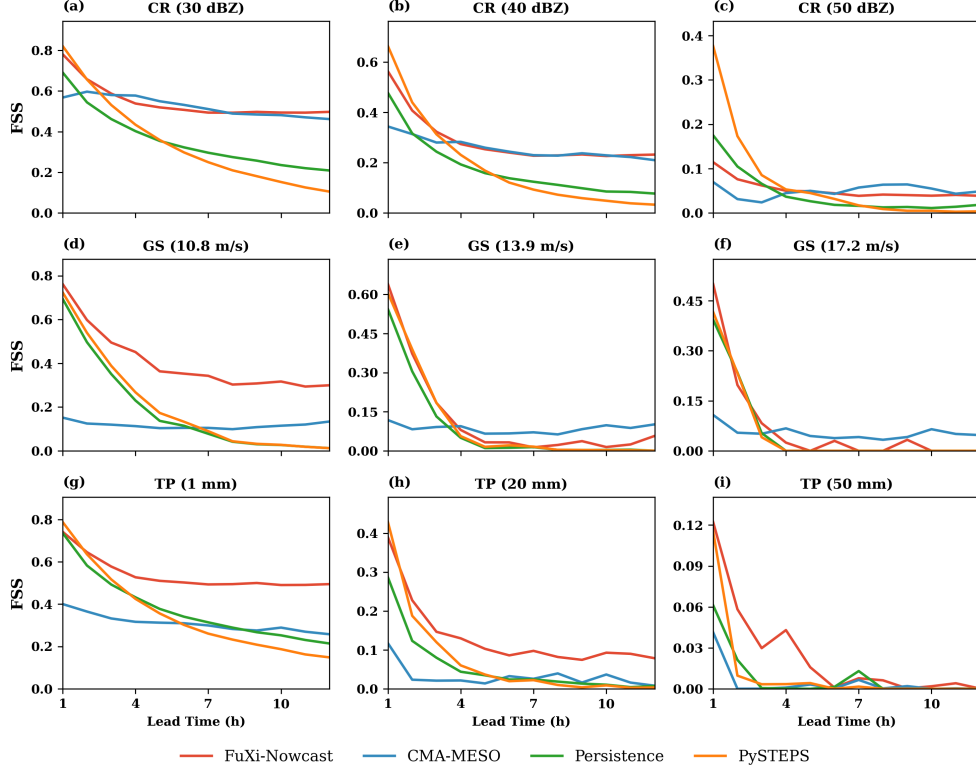
Supplementary Figure 4: Probability of detection across forecast methods, variables, and thresholds. The probability of detection (POD) is compared for FuXi-Nowcast (red), CMA-MESO (blue), Persistence (green), and PySTEPS (orange) for 12 h forecasts. Panels (a)–(c), (d)–(f), and (g)–(i) correspond to composite reflectivity (CR), wind gust (GS), and total precipitation (TP), respectively. Columns show CR thresholds of 30, 40, and 50 dBZ; GS thresholds of 10.8, 13.9, and 17.2 m s^{-1} ; and TP thresholds of 1, 20, and 50 mm. Higher POD indicates that a larger fraction of observed events is detected, and a value of 1 indicates that all observed events are detected.



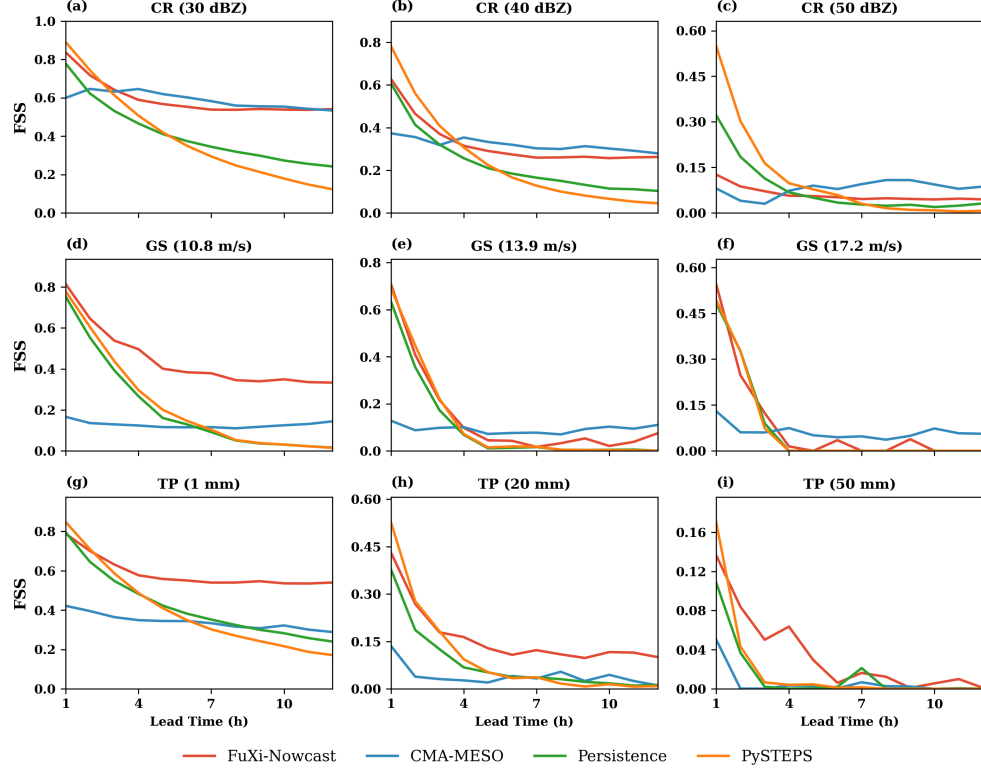
Supplementary Figure 5: Fractions Skill Score at an approximately 5 km neighborhood scale. The Fractions Skill Score (FSS) is compared for FuXi-Nowcast (red), CMA-MESO (blue), Persistence (green), and PySTEPS (orange). Panels (a)–(c), (d)–(f), and (g)–(i) correspond to composite reflectivity (CR), wind gust (GS), and total precipitation (TP), respectively. Columns show CR thresholds of 30, 40, and 50 dBZ; GS thresholds of 10.8, 13.9, and 17.2 m s^{-1} ; and TP thresholds of 1, 20, and 50 mm. Higher FSS indicates better spatial agreement within the neighborhood, and a value of 1 indicates a perfect forecast.



Supplementary Figure 6: Fractions Skill Score at an approximately 10 km neighborhood scale. The Fractions Skill Score (FSS) is compared for FuXi-Nowcast (red), CMA-MESO (blue), Persistence (green), and PySTEPS (orange). Panels (a)–(c), (d)–(f), and (g)–(i) correspond to composite reflectivity (CR), wind gust (GS), and total precipitation (TP), respectively. Columns show the same CR, GS, and TP thresholds as in Supplementary Figure 5. Higher FSS indicates better spatial agreement within the neighborhood, and a value of 1 indicates a perfect forecast.



Supplementary Figure 7: Fractions Skill Score at an approximately 20 km neighborhood scale. The Fractions Skill Score (FSS) is compared for FuXi-Nowcast (red), CMA-MESO (blue), Persistence (green), and PySTEPS (orange). Panels (a)–(c), (d)–(f), and (g)–(i) correspond to composite reflectivity (CR), wind gust (GS), and total precipitation (TP), respectively. Columns show the same CR, GS, and TP thresholds as in Supplementary Figure 5. Higher FSS indicates better spatial agreement within the neighborhood, and a value of 1 indicates a perfect forecast.



Supplementary Figure 8: Fractions Skill Score at an approximately 40 km neighborhood scale. The Fractions Skill Score (FSS) is compared for FuXi-Nowcast (red), CMA-MESO (blue), Persistence (green), and PySTEPS (orange). Panels (a)–(c), (d)–(f), and (g)–(i) correspond to composite reflectivity (CR), wind gust (GS), and total precipitation (TP), respectively. Columns show the same CR, GS, and TP thresholds as in Supplementary Figure 5. Higher FSS indicates better spatial agreement within the neighborhood, and a value of 1 indicates a perfect forecast.

68 **Data sources**

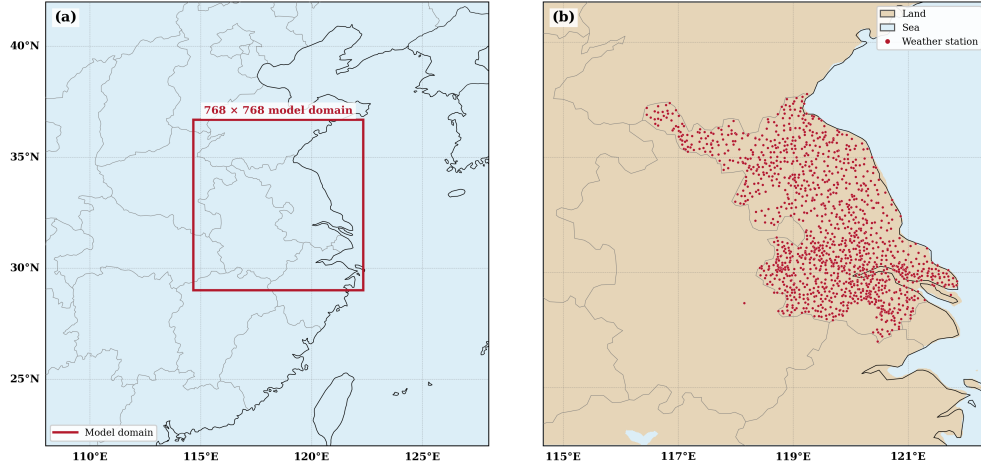
69 **Dataset overview**

Supplementary Table 1: Meteorological datasets used by FuXi-Nowcast, including original and target resolutions, variables and temporal coverage. The Role column indicates whether each variable is used as both input and output or only as input.

Data Source	Role	Raw Resolution	Processed Resolution	Variables	Steps
FuXi-2.0 Forecasts	I	721×1440 (0.25°)	768×768 (0.01°)	5 upper-air variables (65 channels) (Z, T, U, V, Q) at 13 pressure levels (50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, and 1000 hPa) 5 surface variables (T2M, U10M, V10M, MSL, TP)	3
Weather Stations	I & O	— (0°)	768×768 (0.01°)	WS10M, GS	3
Radar Reflectivity	I & O	1100×1100 (0.01°)	768×768 (0.01°)	CR	3
HRCLDAS	I & O	4500×7000 (0.01°)	768×768 (0.01°)	Q2M, TP, U10M, V10M, T2M	3
Land-Sea Mask	I	721×1440 (0.25°)	768×768 (0.01°)	Mask	1

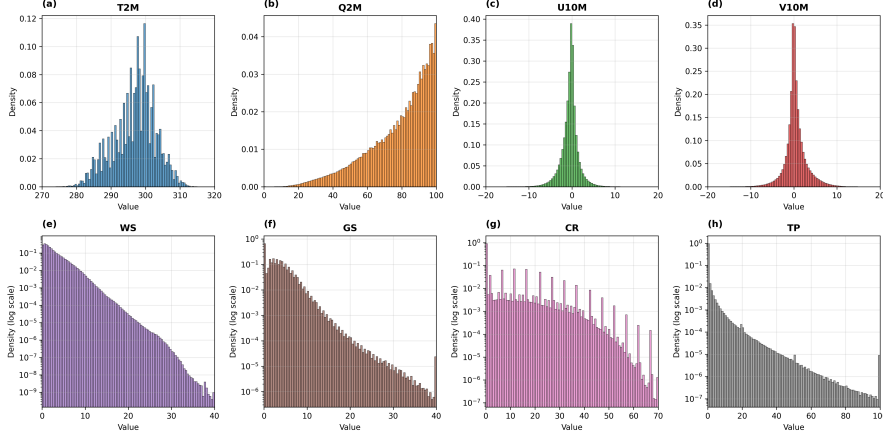
Study domain and station network

Supplementary Figure 9 shows the study domain and the spatial distribution of ground-based weather stations used in this study. The domain consists of 768×768 grid points covering East China (29.01°N - 36.68°N , 114.67°E - 122.34°E) at a spatial resolution of 0.01° . More than 1300 ground-based weather stations within the domain provide hourly observations of wind speed, wind gusts, temperature, and humidity. Occasional data gaps at individual stations do not substantially affect the overall spatial coverage. The station mask used in the evaluation follows this station distribution and is applied only to select grid points for GS verification. CR and TP are evaluated over the entire gridded domain.

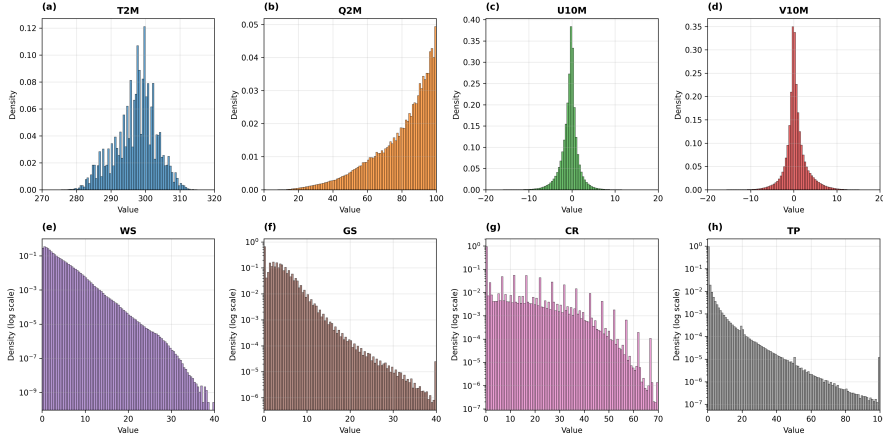


Supplementary Figure 9: Study domain and ground-based weather station network. Panel (a) shows the regional context and the East China model domain, marked by a red box, spanning 29.01–36.68 degrees north and 114.67–122.34 degrees east on a 768×768 grid at 0.01 degree spatial resolution. Panel (b) shows the land–sea mask and the locations of the more than 1300 ground-based weather stations used by FuXi-Nowcast. Red markers indicate station locations represented on the model grid.

80 Training sample distributions



Supplementary Figure 10: Distributions of the eight target forecast variables in the training set before sample filtering. Panels (a)–(h) show 2-m air temperature (T2M; blue bars), 2-m specific humidity (Q2M; orange bars), the 10-m zonal wind component (U10M; green bars), the 10-m meridional wind component (V10M; red bars), wind speed (WS; purple bars), wind gust (GS; brown bars), composite reflectivity (CR; pink bars), and total precipitation (TP; grey bars), respectively. The vertical axis is linear for panels (a)–(d) and logarithmic for panels (e)–(h).



Supplementary Figure 11: Distributions of the eight target forecast variables in the filtered training set. Panels (a)–(h) show 2-m air temperature (T2M; blue bars), 2-m specific humidity (Q2M; orange bars), the 10-m zonal wind component (U10M; green bars), the 10-m meridional wind component (V10M; red bars), wind speed (WS; purple bars), wind gust (GS; brown bars), composite reflectivity (CR; pink bars), and total precipitation (TP; grey bars), respectively. The filtering procedure retains cases with significant precipitation or wind-gust events in the training dataset. The axis conventions are the same as in Supplementary Figure 10.

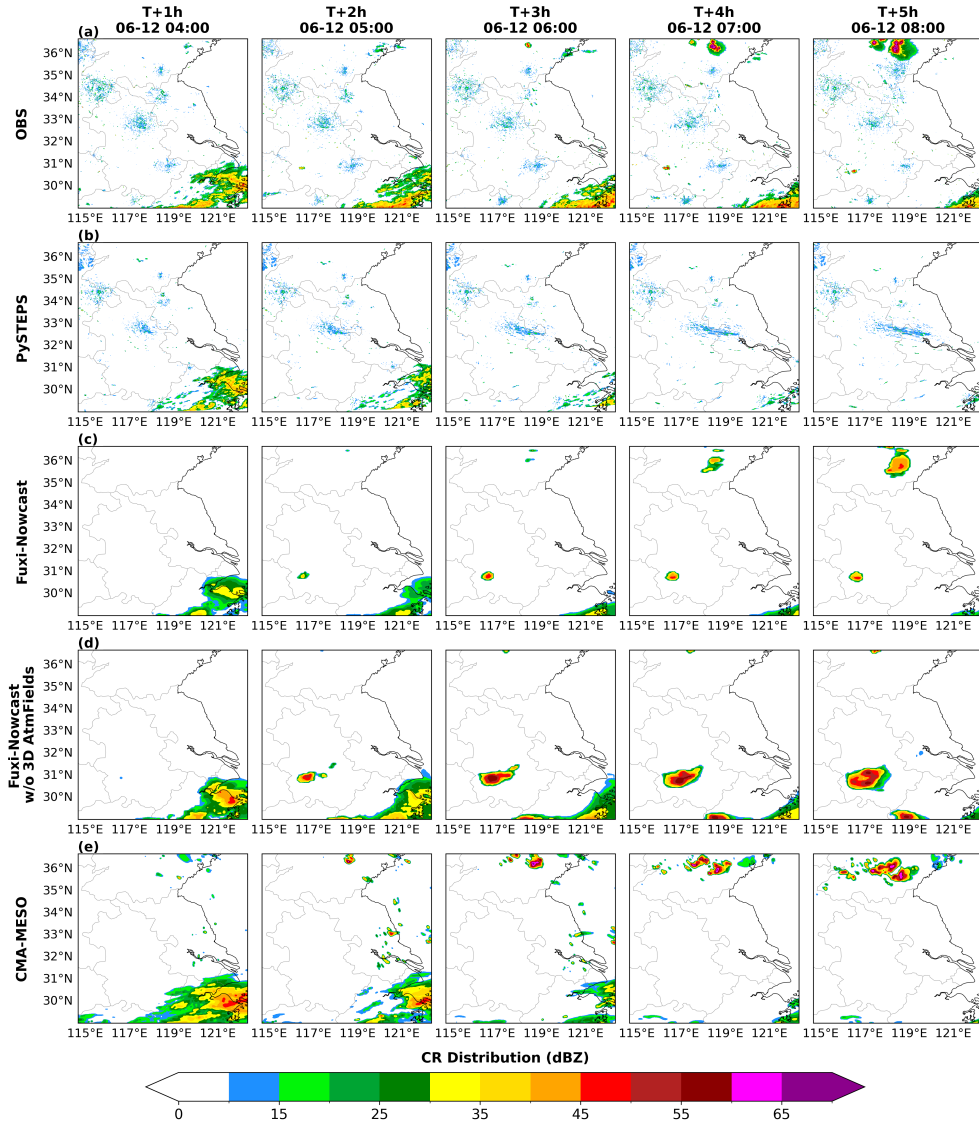
81 Additional convective event cases

82 In addition to the dryline-triggered convective initiation (CI) case presented in the
83 main text (16 June 2025), we examine additional cases from the 2024 evaluation
84 period and the 2025 severe-weather season to show different aspects of FuXi-Nowcast’s
85 convective forecasting behavior, including scattered convective initiation, maintenance
86 of organized convection over extended lead times, and the environmental conditions
87 associated with a squall line and warm-sector convection. These cases are included to
88 cover several typical modes of convective evolution discussed in this study. The 2024
89 cases are drawn from the main evaluation period, whereas the 2025 cases are presented
90 as additional event analyses.

91 Scattered convective initiation case: 12 June 2024 (03 UTC)

92 A scattered convective initiation event occurred over East China on 12 June 2024
93 (Supplementary Figure 12). Observed CR shows weak echoes in the northwestern part
94 of the domain at $T + 1$ h (04 UTC), with isolated convective echoes developing and
95 intensifying through $T + 2$ – 3 h (05–06 UTC) before dissipating by $T + 4$ – 5 h (07–08
96 UTC). FuXi-Nowcast captures this evolution reasonably well: it predicts the initial
97 appearance of scattered echoes in the northwest and their subsequent weakening. The
98 spatial distribution and intensity of the predicted reflectivity are generally consistent
99 with the observations.

100 PySTEPS, by contrast, fails to capture CI and produces only a diffuse signal that
101 rapidly fades. CMA-MESO generates widespread moderate echoes across the northern
102 part of East China from the outset that bear little resemblance to the observed isolated
103 convective initiation in the northwest. The figure also includes the retrained FuXi-
104 Nowcast variant without the three-dimensional atmospheric fields for comparison. This
105 variant fails to predict CR over the northwestern part of the domain, indicating that
106 the three-dimensional atmospheric fields contribute to the CI prediction.

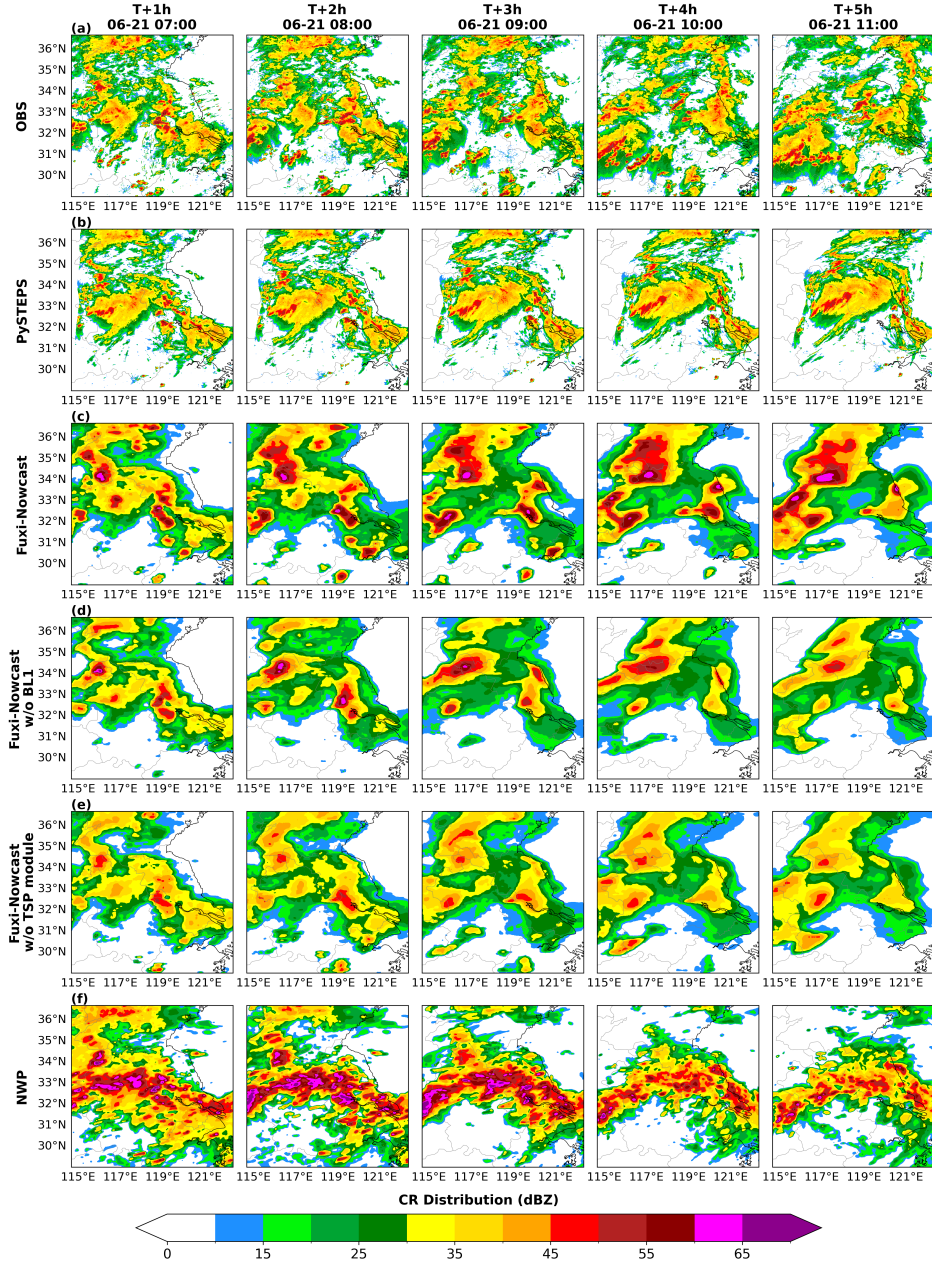


Supplementary Figure 12: Scattered convective initiation case study on 12 June 2024. Composite radar reflectivity (dBZ) over East China for a scattered convective initiation case initialized at 03 UTC 12 June 2024. Rows (a)–(e), from top to bottom, show observations, PySTEPS forecasts, FuXi-Nowcast forecasts, the retrained FuXi-Nowcast variant without the three-dimensional atmospheric fields, and CMA-MESO forecasts, respectively. Columns from left to right correspond to forecast lead times of $T + 1$ to $T + 5$ h, valid from 04 to 08 UTC. Colors denote reflectivity according to the color bar.

Convective maintenance case: 21 June 2024

An extensive convective system affected East China on 21 June 2024 (Supplementary Figure 13). Observed CR at $T + 1$ h (07 UTC) shows a broad area of stratiform and embedded convective precipitation across the western and central parts of the domain. Through $T + 2$ – 5 h (08–11 UTC), the system propagates eastward and intensifies, producing widespread reflectivity exceeding 45 dBZ with embedded convective echoes above 55 dBZ covering most of the province.

FuXi-Nowcast reproduces the large-scale organization and eastward propagation of this convective system. The predicted reflectivity pattern captures both the stratiform background and the embedded convective maxima, while maintaining realistic intensity through the 5-hour forecast window. PySTEPS preserves the initial echo pattern but progressively blurs and weakens the signal with increasing lead time, failing to maintain the convective structure beyond $T + 2$ h. CMA-MESO overestimates the convective intensity and produces overly strong echoes in the southern part of the domain while underrepresenting the precipitation extent in the north. The figure also includes two retrained ablation variants: one without the thresholded signal pooling (TSP) module and one without the Balanced L1 loss (BL1). Relative to the full FuXi-Nowcast, both variants lose the organization and intensity of the convective system more rapidly as lead time increases.



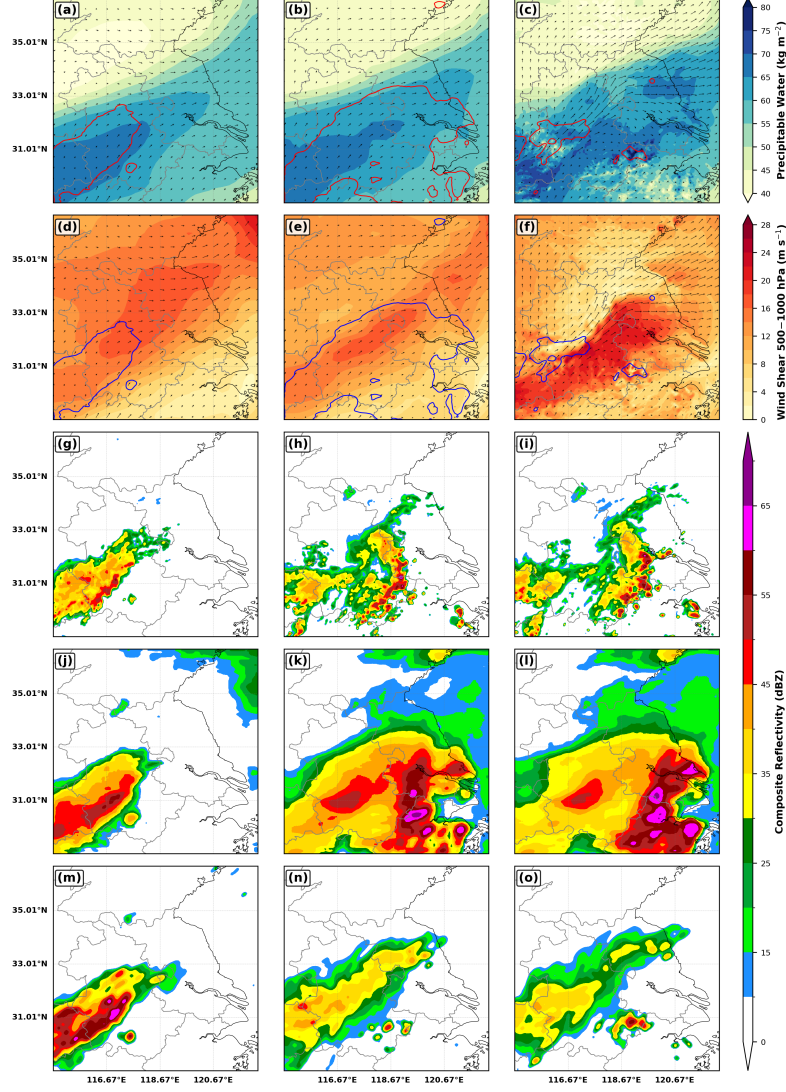
Supplementary Figure 13: Convective maintenance case study on 21 June 2024. Composite radar reflectivity (dBZ) over East China for a convective maintenance case initialized at 06 UTC 21 June 2024. Rows (a)–(f), from top to bottom, show observations, PySTEPS forecasts, FuXi-Nowcast forecasts, two retrained ablation variants without the Balanced L1 loss (BL1) and without the thresholded signal pooling (TSP) module, and CMA-MESO forecasts, respectively. Columns from left to right correspond to forecast lead times of $T + 1$ to $T + 5$ h, valid from 07 to 11 UTC. Colors denote reflectivity according to the color bar.

Squall line case: 27 June 2025

A squall line event occurred over East China on 27 June 2025. Convection had already been triggered at the forecast initialization time (00 UTC). Therefore, we focus on the maintenance and structural evolution of the convective system. Comparison with observed composite reflectivity (CR) from 01 to 12 UTC shows that FuXi-Nowcast generally better reproduces the morphology and position of the squall line than CMA-MESO. The added CR panels in Supplementary Figure 14 provide a direct comparison of observations and forecasts at matched valid times. The observed CR evolves from a compact southwest–northeast-oriented convective band at 01 UTC into a more extended and fragmented system by 05–06 UTC. FuXi-Nowcast retains the elongated organization but produces a broader and more intense echo region, whereas CMA-MESO produces a comparatively weaker and more fragmented structure.

To examine the different structural outcomes in the two models, we compare their environmental fields side by side (Supplementary Figure 14). In the three-dimensional atmospheric fields at $T + 1$ h and $T + 5$ h (panels a, b, d, e), the predicted convective system (delineated by the 35 dBZ contour) is embedded within a region of total precipitable water (TPW) exceeding 70 kg m^{-2} (panels a, b). It is also located within a region where the 500–1000 hPa bulk wind shear exceeds 14 m s^{-1} (panels d, e). The moisture supply is consistent with sustained convective development. The strong deep-layer shear is also consistent with an environment favorable for storm organization and longevity, because it helps separate the updraft from the precipitation-driven downdraft. Together, these factors are consistent with conditions that support squall line formation and maintenance.

In contrast, CMA-MESO at $T + 6$ h (panels c, f) produces TPW as high as 75 kg m^{-2} in the convective region (panel c). However, the 500–1000 hPa wind shear is only about 5 m s^{-1} (panel f), which is less supportive of a coherent linear structure. The convection in CMA-MESO also appears in regions of high moisture content, but it lacks the same degree of dynamical organization seen in the atmospheric field diagnostics. This comparison indicates that wind-shear information may be important for representing this event.



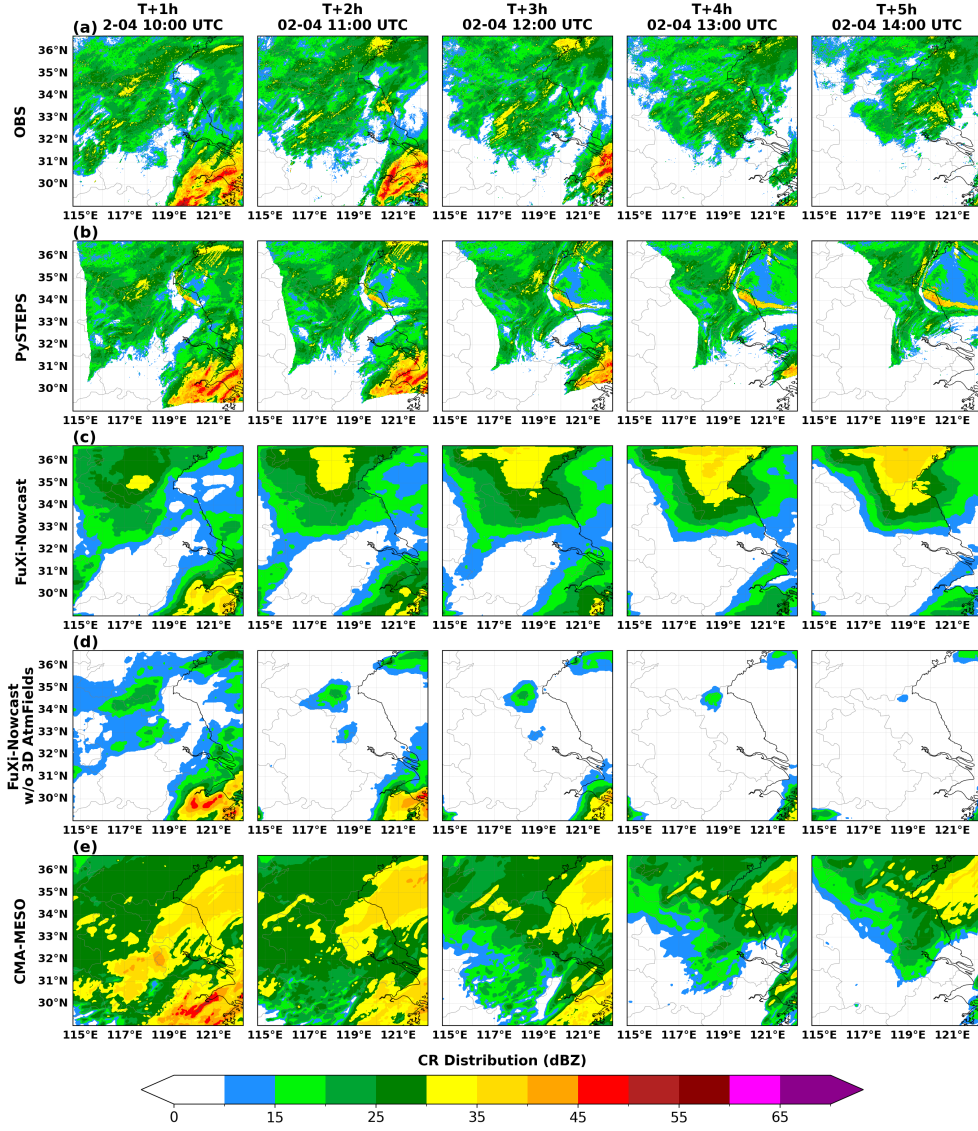
Supplementary Figure 14: Environmental diagnostics and matched composite-reflectivity (CR) fields for the squall-line case on 27 June 2025, initialized at 00 UTC. Panels (a) and (b) show FuXi-2.0 total precipitable water (TPW; shading, kg m^{-2}) valid at 01 and 05 UTC, respectively, and panel (c) shows CMA-MESO TPW valid at 06 UTC; red contours mark 35-dBZ CR from the corresponding FuXi-Nowcast or CMA-MESO forecast, and arrows show 925-hPa winds. Panels (d) and (e) show FuXi-2.0 500–1000-hPa bulk wind-shear magnitude (shading, m s^{-1}) valid at 01 and 05 UTC, and panel (f) shows the corresponding CMA-MESO field valid at 06 UTC; blue contours mark 35-dBZ CR and arrows show 925-hPa winds. Panels (g)–(i), (j)–(l), and (m)–(o) show observed, FuXi-Nowcast, and CMA-MESO CR, respectively, at 01, 05, and 06 UTC from left to right. All shaded CR panels use the same 0–70 dBZ color scale.

Cold season convective case: 4 February 2024

A cold season convective event occurred over East China on 4 February 2024 (Supplementary Figure 15). Observations show pronounced echoes in the southeastern part of the domain at $T + 1$ h (10 UTC), together with weaker and more scattered echoes over the northern part of East China. The convective activity remains evident during the following hours, although its spatial organization and intensity evolve substantially.

FuXi-Nowcast reproduces the broad southeastern convective region and its persistence, but its forecasts are smoother and less spatially detailed than the observations. PySTEPS retains some of the observed structures at the shortest lead times, but the predicted echoes progressively weaken and lose spatial coverage with increasing lead time. CMA-MESO produces widespread and relatively smooth echoes, with differences in the placement and evolution of the main convective areas.

The FuXi-Nowcast variant without the three-dimensional atmospheric fields captures some of the initial southeastern echoes, but its small-scale structures and higher-reflectivity regions decay rapidly with lead time. This qualitative example illustrates the contribution of the three-dimensional atmospheric fields to maintaining coherent convective structures during the forecast period.



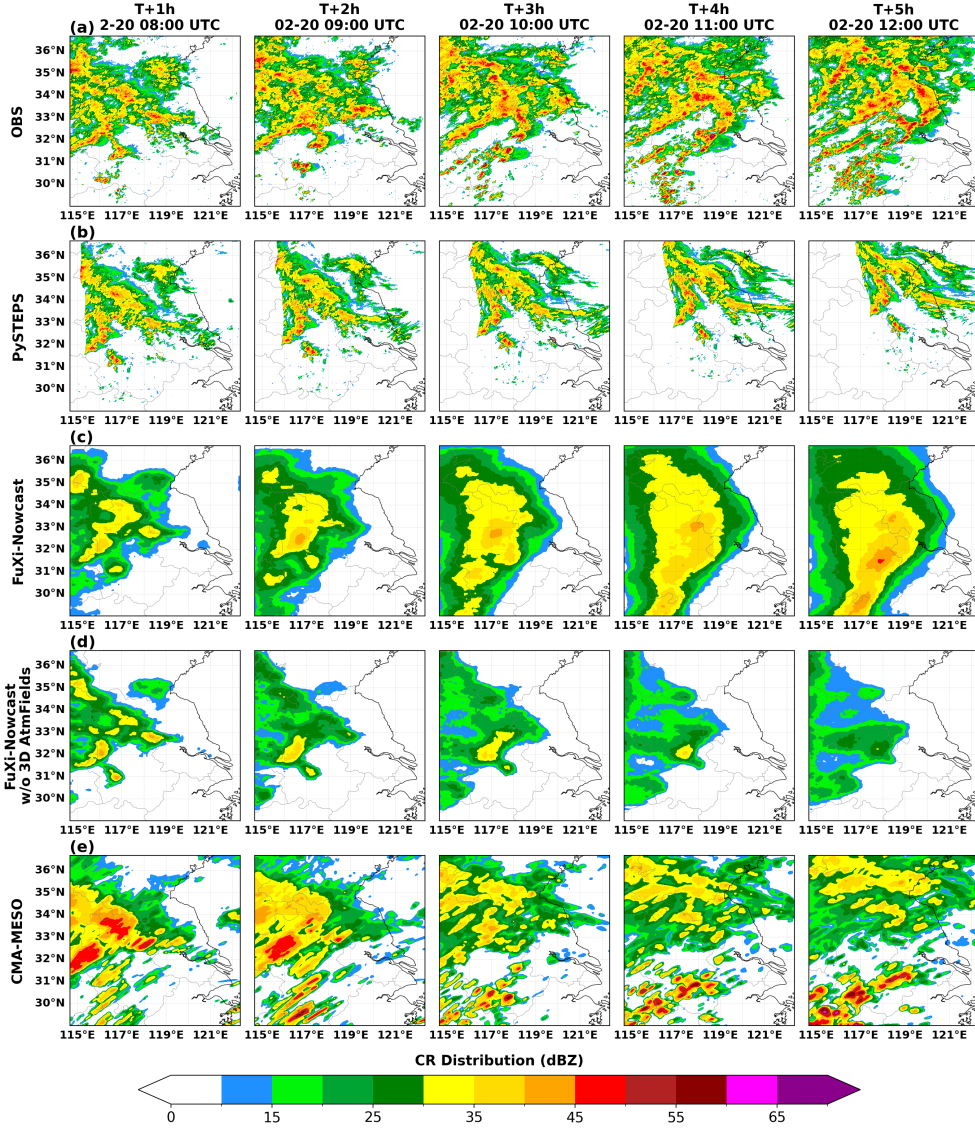
Supplementary Figure 15: Cold season convective case study on 4 February 2024. Composite radar reflectivity (dBZ) over East China for a cold season convective case initialized at 09 UTC on 4 February 2024. Rows (a)–(e), from top to bottom, show observations, PySTEPS forecasts, FuXi-Nowcast forecasts, the retrained FuXi-Nowcast variant without the three-dimensional atmospheric fields, and CMA-MESO forecasts, respectively. Columns from left to right correspond to forecast lead times of $T + 1$ to $T + 5$ h, valid from 10 to 14 UTC. Colors denote reflectivity according to the color bar.

173 Cold season convective case: 20 February 2024

174 An organized cold season convective event occurred over East China on 20 February
175 2024 (Supplementary Figure 16). At $T + 1$ h (08 UTC), the observations show a broad
176 convective system extending across the western and northern parts of the domain. The
177 system subsequently evolves and extends eastward through $T + 5$ h (12 UTC), while
178 retaining substantial spatial variability.

179 FuXi-Nowcast captures the broad scale evolution of the convective system,
180 although the predicted echoes are smoother and contain less small scale variability
181 than the observations. PySTEPS provides a relatively similar spatial structure at the
182 shortest lead times, but its echoes gradually weaken as the lead time increases. CMA-
183 MESO produces widespread convective echoes with a different spatial distribution and
184 evolution from the observations.

185 The FuXi-Nowcast variant without the three-dimensional atmospheric fields repro-
186 duces part of the initial convective structure, but the system weakens substantially
187 and loses spatial extent at later lead times.



Supplementary Figure 16: Cold season convective case study on 20 February 2024. Composite radar reflectivity (dBZ) over East China for a cold season convective case initialized at 07 UTC on 20 February 2024. Rows (a)–(e), from top to bottom, show observations, PySTEPS forecasts, FuXi-Nowcast forecasts, the retrained FuXi-Nowcast variant without the three-dimensional atmospheric fields, and CMA-MESO forecasts, respectively. Columns from left to right correspond to forecast lead times of $T + 1$ to $T + 5$ h, valid from 08 to 12 UTC. Colors denote reflectivity according to the color bar.

Additional ablation results

Supplementary Table 2: Overview of retrained ablation experiment settings for FuXi-Nowcast. Rows list the full model and the retrained ablation variants used in the manuscript. Columns indicate whether each component is retained or removed.

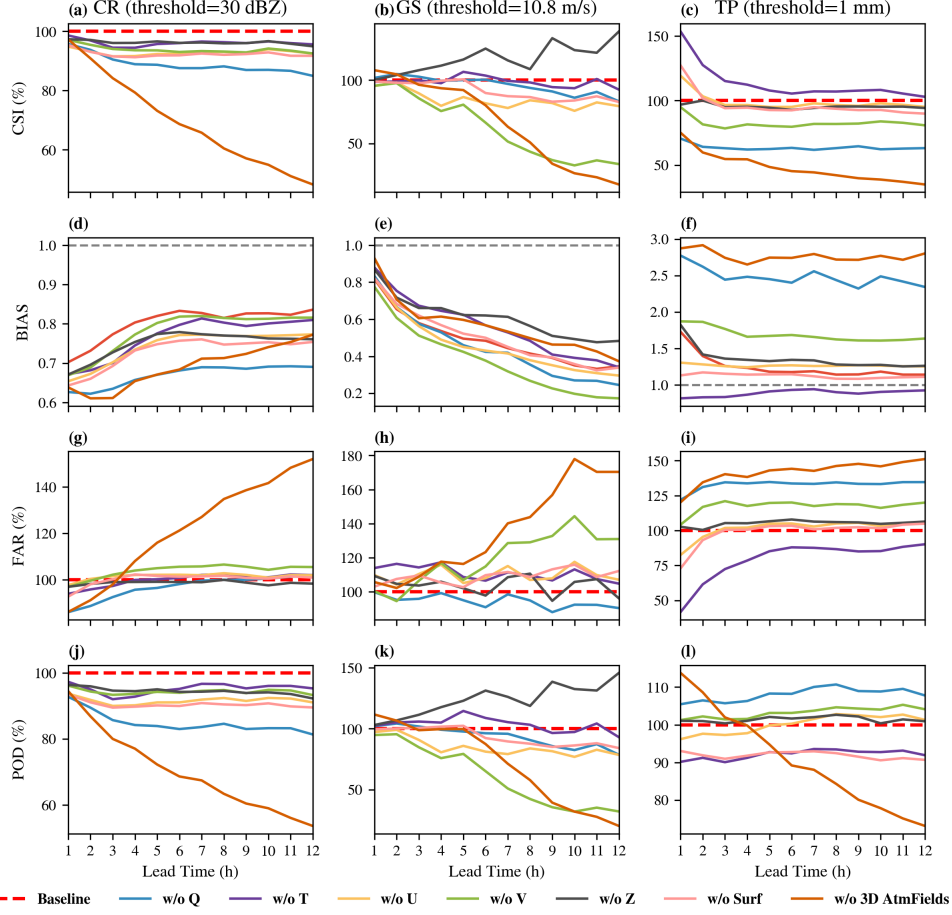
Setting	3D AtmFields	TSP	ASF	BL1	Dice	Focal	LPIPS	SSIM
Full model	✓	✓	✓	✓	✓	✓	✓	✓
w/o 3D atmospheric fields	×	✓	✓	✓	✓	✓	✓	✓
w/o TSP module	✓	×	✓	✓	✓	✓	✓	✓
w/o ASF module	✓	✓	×	✓	✓	✓	✓	✓
w/o BL1	✓	✓	✓	×	✓	✓	✓	✓
w/o Dice	✓	✓	✓	✓	×	✓	✓	✓
w/o Focal	✓	✓	✓	✓	✓	×	✓	✓
w/o LPIPS	✓	✓	✓	✓	✓	✓	×	✓
w/o SSIM	✓	✓	✓	✓	✓	✓	✓	×

Input variable ablation

190 The corrected results show target-dependent sensitivity. Removing Q causes the clear-
 191 est degradation for CR and TP, with relative CSI decreasing to approximately 85%
 192 and 63% of the baseline at 12 h, respectively. Removing the surface variables produces
 193 moderate reductions in CR and TP, to about 92% and 90% of the baseline. GS is most
 194 sensitive to the removal of V, with the 12-h relative CSI decreasing to approximately
 195 34%, followed by U and Surf, which decrease to about 80% and 83%, respectively.
 196 Removing T or Z generally has smaller effects on CR, although the responses vary
 197 across targets and lead times. The observation-only w/o 3D AtmFields configuration
 198 shows substantially larger degradation, with 12-h relative CSI values of approximately
 199 48%, 18%, and 35% for CR, GS, and TP, respectively. These results indicate that both
 200 upper-air moisture and near-surface information contribute to the model predictions.

Supplementary Table 3: Overview of input-variable ablation settings for FuXi-Nowcast. A checkmark indicates that the corresponding input channels are retained, whereas a cross indicates that they are removed or not used. The upper-air groups Z, T, U, V, and Q each contain 13 pressure levels. T2M, U10M, V10M, MSL, and TP denote the five surface atmospheric fields. The eight observation channels are t2m, q2m, u10m, v10m, ws, gs, cr, and tp.

Setting	Upper-air fields					Surface fields					8 observation channels
	Z	T	U	V	Q	T2M	U10M	V10M	MSL	TP	
Full model	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
w/o 3D AtmFields	×	×	×	×	×	×	×	×	×	×	✓
w/o Q	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓
w/o T	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓
w/o U	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓
w/o V	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓
w/o Z	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
w/o Surf	✓	✓	✓	✓	✓	×	×	×	×	×	✓

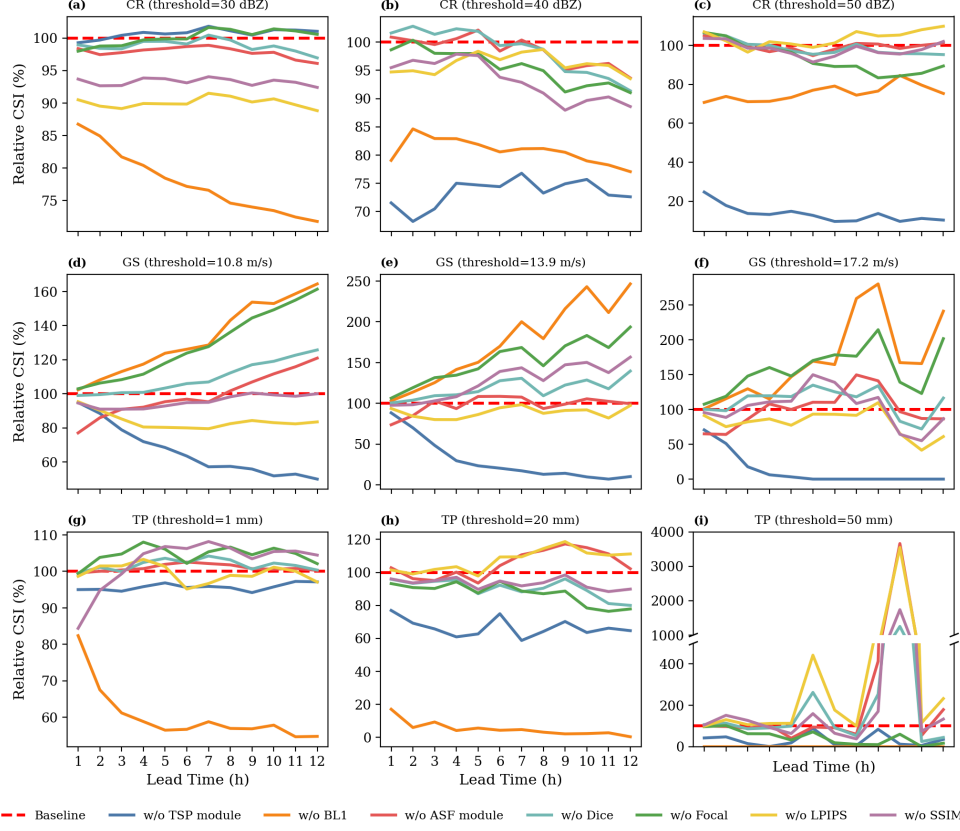


Supplementary Figure 17: Ablation study showing relative verification changes when individual variable groups are removed from the input. Columns correspond to composite reflectivity (CR, threshold = 30 dBZ), wind gust (GS, threshold = 10.8 m/s), and total precipitation (TP, threshold = 1 mm), respectively. Panels (a)–(c), (d)–(f), (g)–(i), and (j)–(l) show relative critical success index (CSI; %), frequency bias (BIAS), relative false-alarm ratio (FAR; %), and relative probability of detection (POD; %), respectively. The red dashed line in the relative CSI, FAR, and POD panels denotes the full-model baseline (100%); in the BIAS panels, the solid red line denotes the full-model baseline and the grey dashed line denotes the perfect value of 1.0. The blue, purple, yellow, green, dark-grey, and pink solid lines denote removal of specific humidity (Q), temperature (T), the zonal wind component (U), the meridional wind component (V), geopotential (Z), and the five surface atmospheric variables, respectively. Those surface variables are 2-m air temperature (T2M), 10-m zonal wind (U10M), 10-m meridional wind (V10M), mean sea-level pressure (MSL), and TP. The orange solid line denotes the observation-only configuration without the three-dimensional atmospheric fields. In the legend, w/o means without, Surf means surface variables, and 3D AtmFields means three-dimensional atmospheric fields.

Model component ablation

The TSP module and the Balanced L1 loss are the two most critical components. Removing the TSP module reduces CSI to $\sim 70\%$ for CR at 30 dBZ and below 20% at 50 dBZ; for GS the degradation is even more severe, with CSI dropping near zero at 17.2 m/s. Removing Balanced L1 causes near-complete failure for heavy precipitation (TP at 20 mm approaching zero) and also degrades CR prediction. The remaining components, such as the Focal loss, Dice loss, learned perceptual image patch similarity (LPIPS) loss, structural similarity index measure (SSIM) loss, and the adaptive signal fusion (ASF) module, each contribute moderate improvements of 5–15% CSI across different variables and thresholds.

These results indicate that the two dominant components act on different aspects of forecast quality. The TSP module is most important for retaining strong convective signatures, while the Balanced L1 loss is most important for rare heavy-precipitation events.



Supplementary Figure 18: Ablation study on model components, showing relative critical success index (CSI; %) as a function of lead time. Panels (a)–(c), (d)–(f), and (g)–(i) correspond to composite reflectivity (CR; 30, 40, and 50 dBZ), wind gust (GS; 10.8, 13.9, and 17.2 m/s), and total precipitation (TP; 1, 20, and 50 mm), respectively. The red dashed line denotes the baseline performance (100%). The blue, orange, red, teal, green, yellow, and purple solid lines denote removal of the thresholded signal pooling (TSP) module, Balanced L1 (BL1) loss, adaptive signal fusion (ASF) module, Dice loss, Focal loss, learned perceptual image patch similarity (LPIPS) loss, and structural similarity index (SSIM) loss, respectively. The TP 50 mm panel (i) uses a broken vertical axis to show several extreme relative values.

References