

<https://doi.org/10.1038/s41612-026-01419-3>

Complementary error structures of AI and numerical models in forecasting boundary-layer jets over the South China Sea

Haimin Li^{1,2,3} & Yu Du^{1,2,3} ✉

The boundary-layer jet (BLJ) over the northern South China Sea exerts a strong influence on heavy rainfall along the South China coast, yet its predictability across different modeling frameworks remains insufficiently understood. This study presents a systematic comparison of BLJ forecasts from a conventional numerical weather prediction (NWP) model (the European Centre for Medium-Range Weather Forecasts, ECMWF) and a data-driven artificial intelligence (AI) model (Pangu). The AI model outperforms the NWP model in overall accuracy, with reduced root-mean-square errors and improved representation of wind direction. However, the two models exhibit distinct and opposite systematic error structures. ECMWF consistently overestimates BLJ intensity and spatial extent, accompanied by a northwestward displacement of the jet core and a clockwise wind-direction bias. Conversely, Pangu underestimates BLJ intensity, with a northeastward core shift and a counterclockwise wind-direction bias. Motivated by these complementary errors, we develop a U-Net-based blending framework with a composite loss function integrating mean absolute error, miss rate, and false alarm rate. The blended forecasts significantly outperform both individual models across multiple BLJ metrics. These results demonstrate that combining physics-based and AI-based forecasts can mitigate systematic biases to improve BLJ prediction, highlighting a promising pathway toward hybrid forecasting systems for weather extremes.

Boundary-layer jets (BLJs), a class of low-level jets (LLJs), are a ubiquitous global phenomenon observed across every continent¹. Characterized by wind speed maxima below approximately 1 km, strong vertical shear, and pronounced diurnal variability, BLJs are shaped by interactions among boundary-layer processes, synoptic or subsynoptic systems, large-scale circulations, and regional topography^{2–4}. They are particularly prevalent in regions influenced by major mountain ranges or significant land-sea thermal gradients^{1,5}. A prominent example is the U.S. Great Plains low-level jet (GPLLJ), which serves as a classic case for investigating jet dynamics and their hydrological impacts due to the availability of extensive observational networks^{4,6–8}.

Beyond continental BLJs, coastal BLJs have also been extensively explored because of their influence on coastal weather and precipitation patterns^{9–12}. Over the northern South China Sea (SCS), BLJs play a critical

role in regulating both the intensity and spatial distribution of coastal heavy rainfall in South China^{2,3,13–15}. They transport large amounts of warm, moist air toward the coast and generate intense low-level convergence near their terminus, thereby creating thermodynamically favorable conditions for deep convection^{13,16,17}. Statistical analyses further show that most coastal warm-sector heavy rainfall events, which are particularly difficult to predict, are closely associated with BLJs, as their spatial structure and temporal evolution exert strong control on rainfall initiation and development^{2,3,10,12,14,18–20}.

Given their central role in coastal heavy rainfall, marine BLJs serve as a key precursor for predicting extreme precipitation over South China, particularly challenging coastal warm-sector rainfall. Numerical modeling studies indicate that rainfall forecast skill in these events is highly sensitive to BLJ intensity and wind direction, and depends strongly on boundary-layer

¹School of Atmospheric Sciences, Sun Yat-sen University, and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China.

²Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, Sun Yat-sen University, Zhuhai, China. ³Key Laboratory of Tropical Atmosphere-Ocean System, Sun Yat-sen University, Ministry of Education, Zhuhai, China. ✉e-mail: duyu7@mail.sysu.edu.cn

parameterization schemes¹⁵. Assimilation experiments further demonstrate that incorporating wind profiles near the marine BLJ can substantially improve rainfall forecasts over South China, highlighting the importance of accurately representing marine BLJs in operational prediction systems²¹.

Numerical weather prediction (NWP) remains the dominant approach for operational weather forecasting, with steady improvements driven by advances in physics-based modeling and computational power²². Nevertheless, models still exhibit persistent deficiencies in simulating boundary-layer jets. For example, high-resolution Weather Research and Forecasting (WRF) simulations tend to underestimate low-level-jet-related coastal heavy rainfall, primarily due to overestimated BLJ wind speeds and a northward displacement of the offshore jet over the SCS²⁰. These errors weaken coastal wind speed gradients and low-level convergence, ultimately reducing simulated coastal precipitation intensity. Similarly, ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) also exhibit substantial spread in BLJ intensity and direction, with differences of approximately 3 m s^{-1} between good and poor ensemble members leading to significant discrepancies in coastal convergence².

In parallel, artificial intelligence (AI)-based weather forecasting models have advanced rapidly in recent years and now rival conventional NWP systems in many aspects^{23–26}. Deep-learning architectures trained on large reanalysis datasets have demonstrated remarkable success in medium-range weather forecasting^{23,27,28}. One prominent example is the Pangu-Weather forecasting system (hereafter Pangu), which demonstrates high forecast accuracy and exceptional computational efficiency²⁹. While Pangu performs well for large-scale circulation features (e.g., geopotential height at 500 hPa)

and tropical cyclone tracks, its capability in representing BLJs, particularly marine BLJs, remains largely unexplored.

More fundamentally, existing evaluations of AI-based forecasts have primarily relied on bulk statistical metrics such as root-mean-square error (RMSE) or anomaly correlation coefficient (ACC). These metrics, while useful, do not adequately capture physically meaningful characteristics of BLJs, including their spatial extent, intensity, jet-core location, and wind direction. As a result, how AI-based and physics-based models differ in representing BLJs—and whether their forecast errors exhibit systematic or complementary structures—remains an open and largely unaddressed question.

In this study, we systematically evaluate and compare the forecast performance of BLJs over the northern SCS between a conventional NWP model (ECMWF) and an AI-based forecasting system (Pangu). Specifically, we address two key scientific questions: (1) What are the respective strengths and limitations of NWP and AI models in forecasting BLJs and their key characteristics? (2) Can a deep-learning-based blending approach effectively exploit complementary error structures to improve BLJ prediction?

Results

Comparison of BLJ forecast performance between Pangu and ECMWF

We first evaluate the performance of ECMWF and Pangu in forecasting BLJ occurrence using the threat score (TS) across various lead times. Overall, the two models show comparable TS values, although Pangu achieves higher TS scores at longer lead times (Fig. 1a). Examination of the FAR and MR reveals distinct and systematic contrasts between the two models: ECMWF

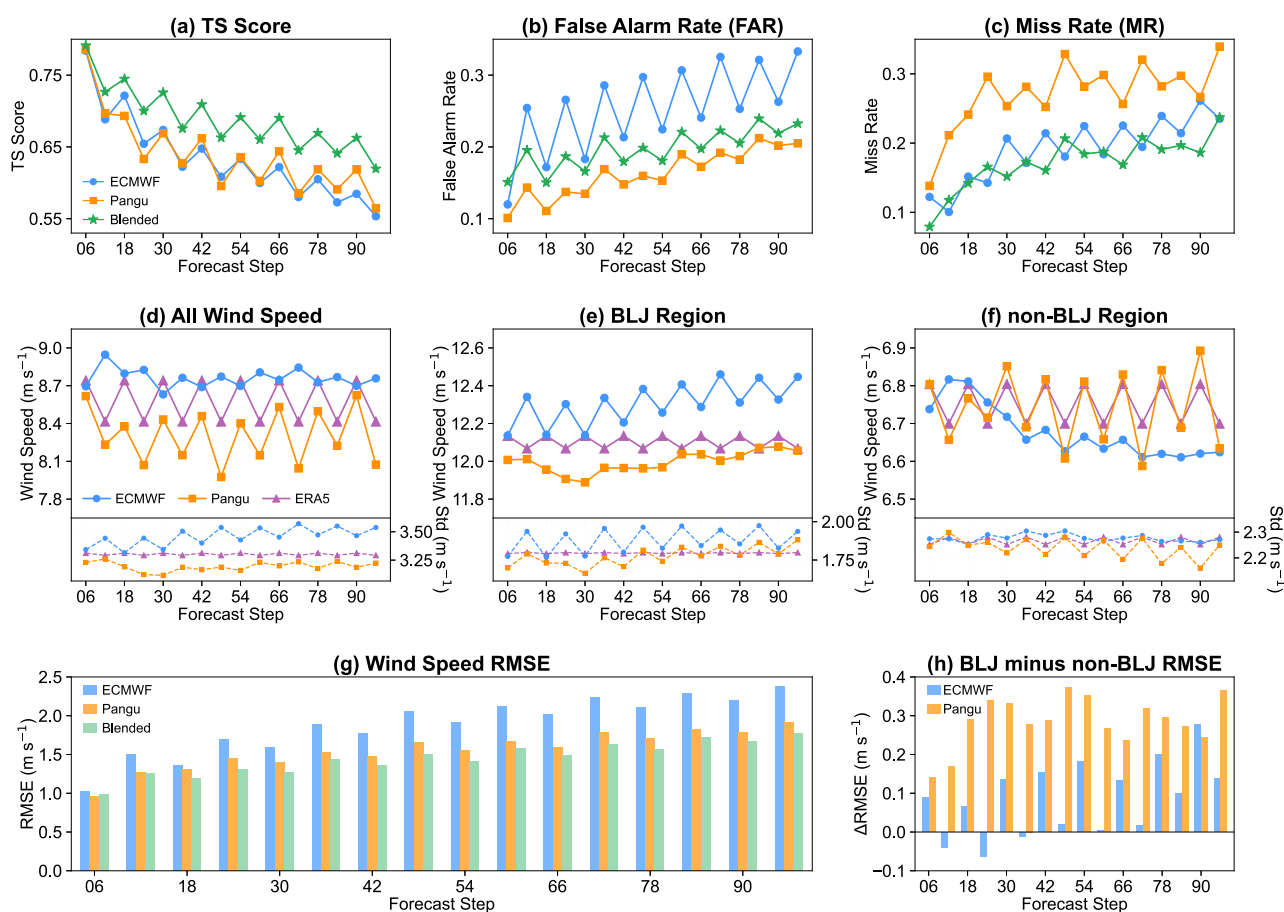


Fig. 1 | Evaluation of model performance in representing BLJ characteristics. **a–c** Threat Score (TS), False Alarm Rate (FAR), and Miss Rate (MR) for BLJ occurrence. Mean (solid lines) and standard deviation (dashed lines) of wind speed for **d** all samples, **e** BLJ events, and **f** non-BLJ events. **g** RMSE (m s^{-1}) for all wind

speed samples. **h** Difference in RMSE (m s^{-1}) between BLJ and non-BLJ events. Orange, blue, and green lines represent forecasts from Pangu, ECMWF, and the U-Net-based blended model, respectively, while purple lines in **(d–f)** denote the ERA5 reanalysis.

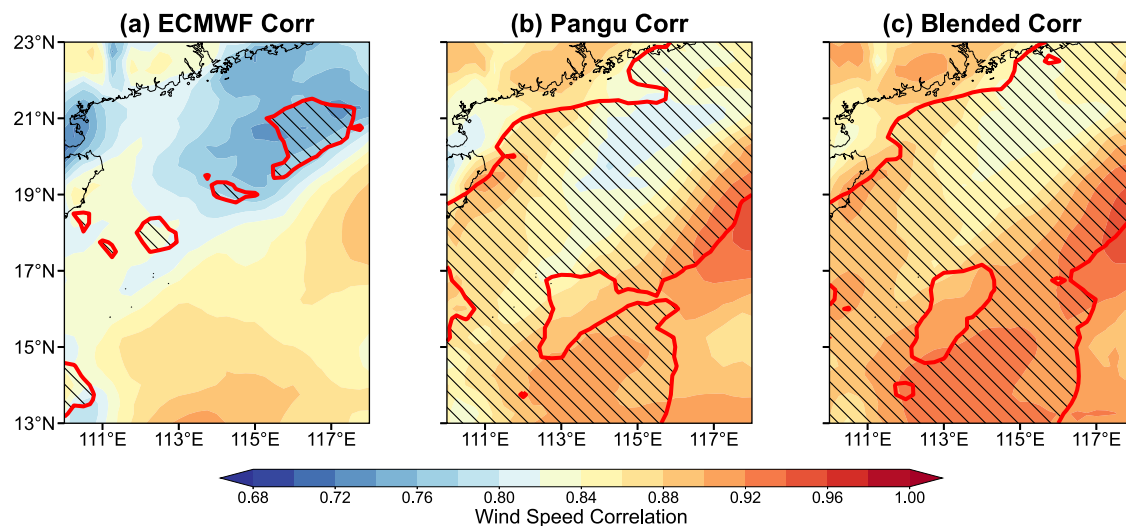


Fig. 2 | Spatial patterns of the correlation coefficient. Shading represents the correlation coefficients of wind speed between ERA5 and **a** ECMWF, **b** Pangu-Weather, and **c** the U-Net-based blended model. Hatched regions denote areas where the correlation coefficient of wind direction exceeds 0.97.

consistently exhibits a higher FAR (Fig. 1b), indicating a tendency to overpredict BLJ occurrences due to its overestimation of low-level wind speeds. Conversely, Pangu shows a higher MR (Fig. 1c), reflecting its systematic underestimation of BLJ intensity and a correspondingly higher frequency of missed BLJ events.

Notably, a clear “seesaw” pattern is observed among TS, FAR, and MR across lead times (Fig. 1a–c). Lead times of 6, 18, ..., and 90 h correspond to BLJ events at 06 and 18 UTC (14:00 and 02:00 LST), while 12, 24, ..., and 96 h correspond to 12 and 00 UTC (20:00 and 08:00 LST). At 02:00 LST, BLJs are primarily driven by relatively regular and predictable diurnal forcing. At 14:00 LST, BLJs tend to be weaker and less frequent, making them comparatively easier for models to predict. In contrast, during transition periods such as sunrise and sunset (around 08:00 and 20:00 LST), BLJ evolution is more strongly influenced by interactions between synoptic-scale processes and rapidly evolving boundary-layer processes. Forecast models tend to perform better under dominant diurnal forcing, but their skill decreases during these transition periods, producing alternating higher and lower detection skill across lead times, which manifests as the observed “seesaw” pattern.

We next assess overall errors in the low-level wind field using the RMSE. Pangu consistently yields lower RMSE than that of ECMWF for all wind speed samples (Fig. 1g), demonstrating its superior overall predictive accuracy. When stratified by event type, however, a clear regime dependence emerges. ECMWF displays comparable RMSE values for BLJ and non-BLJ events, while Pangu achieves significantly lower RMSE during non-BLJ events than during BLJ events (Fig. 1h). This contrast reflects the inherent tendency of deep-learning models to prioritize learning the dominant, near-mean states of the data distribution, at the expense of rarer extreme events.

We further examine the spatial correlation between the forecasts and the ERA5 reanalysis (Fig. 2a, b). The results show that Pangu achieves generally higher correlation coefficients across the study domain than ECMWF, indicating stronger spatial agreement with the reference data. This advantage is evident for both wind speed and wind direction, suggesting a more consistent representation of the low-level wind field, which is consistent with the RMSE-based evaluation.

Consistent with this interpretation, the mean wind speed evolution in Fig. 1d–f shows that Pangu systematically underestimates wind speeds relative to ERA5 across all lead times, with the largest negative bias occurring during BLJ events. In contrast, ECMWF exhibits a clear positive wind speed bias, particularly during BLJ conditions. Analysis of wind speed variability further highlights fundamental differences between the two models. Across all samples, BLJ events, and non-BLJ events, ECMWF exhibits the largest standard deviation in wind speed, followed by ERA5, and then Pangu (Fig.

1d–f). This reduced spread in Pangu forecasts reflects a tendency toward regression to the climatological mean, resulting in smoother wind fields and diminished sensitivity to high-intensity BLJ occurrences. These characteristics help explain the absence of the pronounced “seesaw” pattern in Pangu under BLJ conditions (Fig. 1e): the systematically weaker winds limit its ability to capture the temporal variations seen in ERA5. In contrast, under non-BLJ conditions (Fig. 1f), Pangu more effectively reproduces the temporal fluctuations present in ERA5.

Model performance is further evaluated using BLJ-specific structural indices. Regarding the BLJ area index, Pangu consistently underestimates the spatial extent of the jet, whereas ECMWF tends to overestimate it (Fig. 3a). When averaged over all 16 lead times, ECMWF exhibits smaller absolute errors than Pangu for this index (Table 1). A similar contrast is evident for the BLJ intensity index (Fig. 3b), with systematic underestimation by Pangu and overestimation by ECMWF. The latter bias is consistent with previous findings from high-resolution WRF simulations over the South China Sea²⁰. In contrast, Pangu demonstrates improved accuracy in forecasting BLJ wind direction. Its forecasts exhibit only a counterclockwise bias, whereas ECMWF exhibits a pronounced clockwise deviation (Fig. 3c), consistent with the well-documented veering bias of near-surface winds in ECMWF forecasts over the Northern Hemisphere³⁰. With respect to the BLJ core position, ECMWF’s forecasts show a systematic northwestward displacement relative to ERA5, while Pangu shows a northeastward shift (Fig. 3d), indicating distinct spatial error structures in the two models.

Composite spatial analysis further substantiates these index-based findings (Fig. 4). The composite wind fields confirm that ECMWF systematically produces stronger jets (Fig. 4b) than observed (Fig. 4a), while Pangu generates weaker wind fields (Fig. 4c), consistent with their respective intensity biases. Corresponding wind-direction anomalies (Fig. 4e, f) reveal a coherent clockwise rotation in ECMWF and a prevailing counterclockwise bias in Pangu, providing spatially explicit confirmation of the wind-direction errors diagnosed by the BLJ indices.

A notable feature of ECMWF’s forecast is a pronounced sawtooth-like oscillation in BLJ area and intensity errors, with minima at lead times of 6, 18, ..., and 90 h (corresponding to 14:00 and 02:00 LST) and maxima at 12, 24, ..., and 96 h (20:00 and 08:00 LST), depending on initialization at 00:00 or 12:00 UTC. This oscillatory behavior is absent in Pangu forecasts despite identical initialization, suggesting it originates from deficiencies in model physics rather than from initial condition errors. Indeed, ECMWF exhibits larger errors during the morning and evening transition periods (Fig. 3e, f), consistent with previous studies showing that NWP models often struggle during rapid stability transitions in the

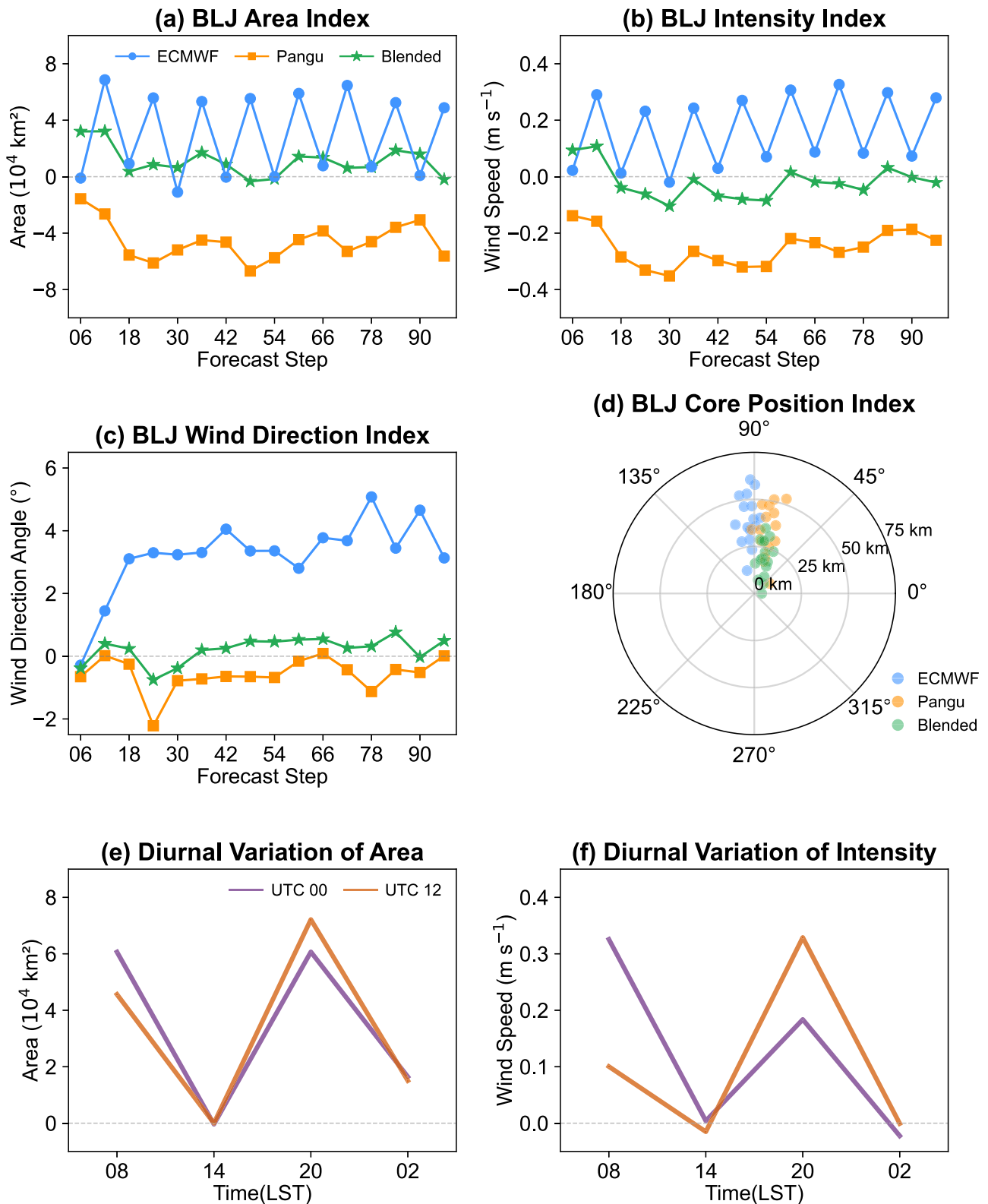


Fig. 3 | Mean errors in BLJ structural indices. Mean errors in **a** BLJ area index (10^4 km^2), **b** BLJ intensity index (m s^{-1}), **c** mean wind direction ($^\circ$, where positive and negative values denote clockwise and counterclockwise biases, respectively), and **d** BLJ core position (km). Orange, blue, and green lines represent forecasts from

Pangu, ECMWF, and the blended model, respectively. Diurnal variations of mean errors in **e** the BLJ area index and **f** the BLJ intensity index. Purple and yellow lines represent forecasts initialized at 00:00 UTC and 12:00 UTC, respectively.

boundary layer³¹. These deficiencies are commonly attributed to limitations in turbulence and planetary boundary-layer (PBL) parameterizations, including insufficient representation of nocturnal stabilization following sunset and delayed establishment of vertical mixing after

sunrise^{32–34}. In contrast, the Pangu model does not rely on explicit physical parameterizations but instead learns statistical relationships directly from data, thereby avoiding systematic, diurnal-cycle-related errors inherent in traditional NWP models.

Table 1 | Performance comparison of the U-Net-based blended forecast against ECMWF and Pangu models

	ECMWF	Pangu	Blended	Skill vs. ECMWF (%)	Skill vs. Pangu (%)
TS	0.634	0.639	0.689	8.68*	7.82*
BLJ RMSE (m s ⁻¹)	1.887	1.556	1.447	23.32*	7.01*
non-BLJ RMSE (m s ⁻¹)	1.752	1.441	1.367	21.97*	5.14*
Area (km ²)	30934.280	45690.557	11950.247	61.37*	73.85*
Intensity (m s ⁻¹)	0.165	0.252	0.051	69.09*	79.76*
Wind direction (°)	3.251	0.586	0.406	87.51*	30.72*
Core position (km)	39.546	34.359	19.702	50.18*	42.66*

All metrics in Table 1 represent averages of absolute mean errors across 16 forecast lead times. ‘Skill vs. ECMWF’ and ‘Skill vs. Pangu’ represent the percentage improvement of the blended forecast relative to ECMWF and Pangu, respectively. BLJ RMSE refers to the RMSE of the low-level wind fields during identified BLJ events, whereas non-BLJ RMSE is computed for wind fields that do not meet the predefined criteria for BLJ events. Statistical significance is assessed using a two-tailed t-test across 16 forecast lead times, with * indicating significance at the 99% confidence level.

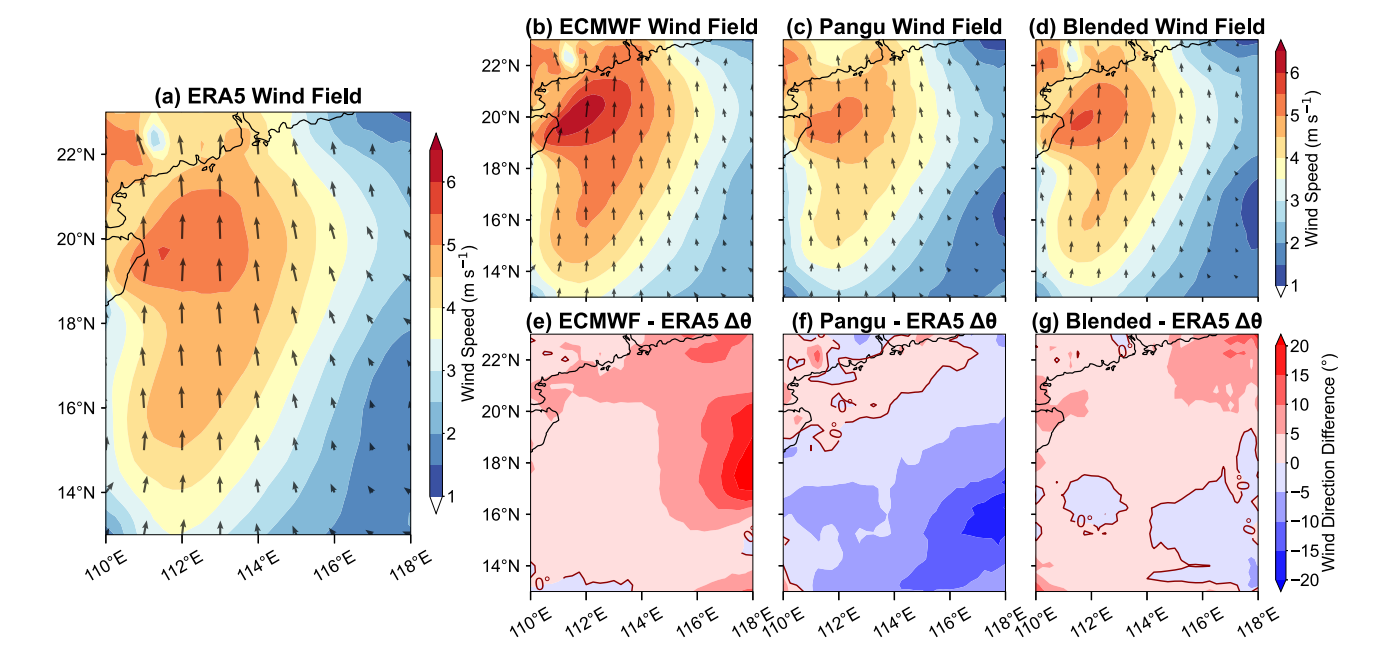


Fig. 4 | Composite analysis of BLJ events across forecast lead times. a–d Composite wind fields at 925 hPa (shaded, m s⁻¹) from ERA5, ECMWF, Pangu, and the U-Net-based blended forecast. **e–g** Wind-direction anomalies relative to ERA5 (shaded, °), with the dark red line denoting the 0° contour.

Blended forecasting by the U-Net model

Building on the preceding analysis, ECMWF and Pangu exhibit systematic and complementary biases in BLJ forecasts, with ECMWF tending to overestimate and Pangu to underestimate BLJ intensity. To exploit these complementary error structures, a U-Net-based blending framework is employed to combine the strengths of the two forecasts. After fine-tuning on the validation set, the blended model is evaluated using independent BLJ events from April to June during 2020–2024. To provide a comprehensive assessment across forecast horizons, the metrics presented in Table 1 are computed by averaging the absolute mean errors across all 16 lead times.

The blended forecast shows a marked improvement in BLJ event detection skill. As illustrated in Fig. 1a, the TS of the blended forecast consistently exceeds that of both ECMWF and Pangu, with the advantage becoming more pronounced at longer lead times. Averaged across all lead times, TS increases by 8.68% relative to ECMWF and by 7.82% relative to Pangu. In addition, the blended forecast achieves a more balanced trade-off between false alarms and missed events. Its FAR is substantially reduced relative to ECMWF, while remaining only slightly higher than that of Pangu (Fig. 1b). More importantly, the MR is markedly lower than that of Pangu and comparable to ECMWF (Fig. 1c), indicating a significant reduction in missed BLJ events without introducing excessive false detections.

Improvements are also evident in overall wind speed accuracy. The blended forecast exhibits a significantly reduced RMSE for all wind speed samples (Fig. 1g), with mean reductions of 23.32% relative to ECMWF and 7.01% relative to Pangu. Notably, these gains are consistent across different conditions. Even under non-BLJ events, the blended model maintains superior performance, with RMSE reductions of 21.97% and 5.14% compared to ECMWF and Pangu, respectively.

In addition, the spatial distribution of correlation coefficients for both wind speed and wind direction (Fig. 2c) shows that the blended model consistently improves agreement with the reference low-level wind field over most regions. Consistent improvements are observed across all BLJ-specific structural indices (Fig. 3). Compared with both standalone models, the blended forecast consistently reduces mean errors in BLJ spatial extent, intensity, wind direction, and jet-core position, as summarized in Table 1.

Composite spatial analyses further support these statistical improvements. The blended wind field shows the closest agreement with ERA5 (Fig. 4a–d), effectively reconciling the overly strong BLJs in ECMWF and the overly weak jets in Pangu. Correspondingly, wind-direction anomalies in the blended forecast are substantially weakened relative to both parent models (Fig. 4g), indicating an effective correction of their opposing directional biases.

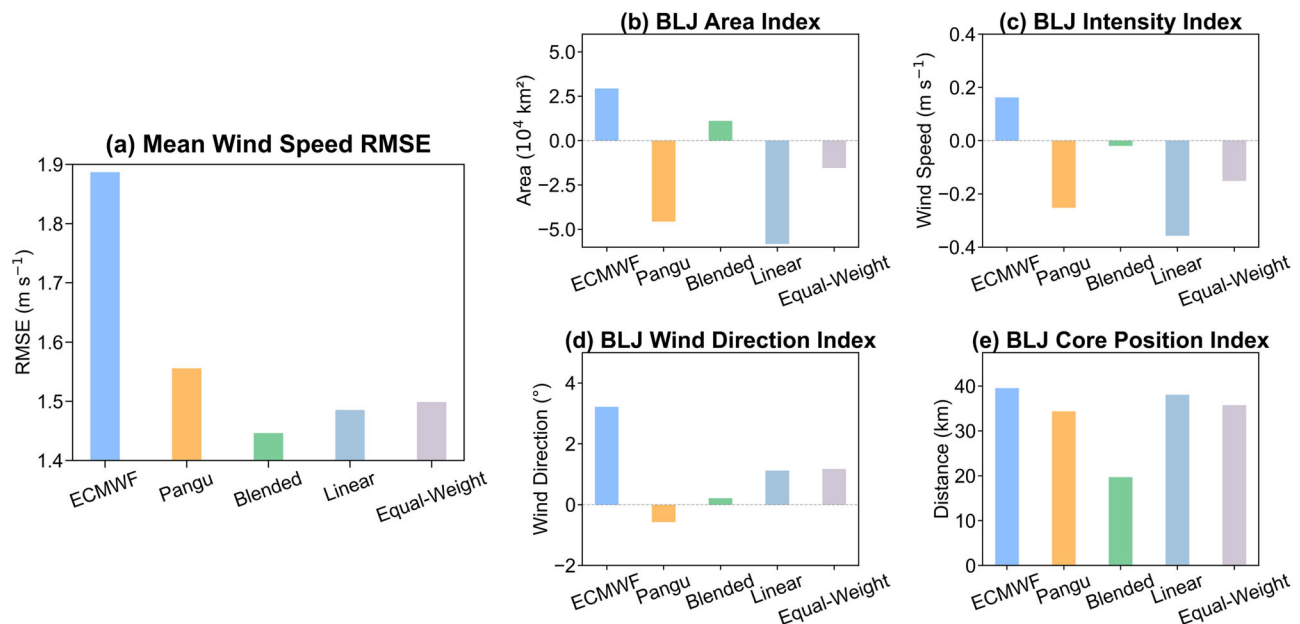


Fig. 5 | Comprehensive evaluation of model performance in the low-level wind field and BLJ characteristics. **a** RMSE (m s^{-1}) for the entire low-level wind field, averaged over 16 forecast lead times. Mean errors in BLJ indices, including **b** BLJ area index (10^4 km^2), **c** BLJ intensity index (m s^{-1}), **d** mean wind direction ($^\circ$), where positive and negative values denote clockwise and counterclockwise biases, respectively), and **e** BLJ core position (km), averaged over 16 forecast lead times.

Orange, blue, and green bars represent forecasts from Pangu, ECMWF, and the U-Net-based blended model, respectively, while light blue and purple bars denote the baseline results from linear regression and an equally weighted average.

Overall, the U-Net-based blending approach effectively exploits the complementary error characteristics of physics-based and AI-based forecasts, yielding more accurate and physically consistent BLJ predictions across multiple indices. To assess whether these improvements stem from the U-Net's nonlinear learning capability rather than simple linear error cancellation between forecasts with opposing biases, we include two baseline statistical approaches for comparison: a linear regression model and an equal-weight averaging method. This comparison isolates the added value of a deep-learning architecture over simpler statistical blending techniques.

The results show that, although the linear regression model reduces overall RMSE relative to ECMWF and Pangu (Fig. 5a), it substantially underestimates BLJ intensity and spatial extent (Fig. 5b, c), with underperforming Pangu in some aspects. Conversely, the equal-weight averaging method improves BLJ-specific metrics, such as area and intensity indices (Fig. 5b, c), but at the cost of a slight degradation in overall wind field accuracy, as evidenced by increased RMSE (Fig. 5a).

The U-Net-based blending model effectively resolves this performance trade-off. By leveraging the complementary error structures of ECMWF and Pangu, it achieves optimal performance across all evaluated aspects, including BLJ intensity, spatial distribution, wind speed, and direction. This demonstrates that the improvement is not merely due to bias compensation, but to the model's ability to dynamically extract and fuse multi-scale features, thereby maintaining global accuracy while accurately representing localized extremes. These results highlight the advantage of deep-learning architectures for complex meteorological forecast fusion.

Interestingly, we further employ SHAP to quantitatively assess the contributions of different input channels within the U-Net-based blending model. The results indicate that approximately 80% of the contribution originates from Pangu-Weather, while ECMWF accounts for about 20%. This ratio is broadly consistent with the weights obtained from the linear regression-based baseline. However, despite this similarity in relative importance, the predictive performance of the linear regression model is significantly inferior to that of the U-Net. This contrast suggests that the proposed blending framework cannot be interpreted as a simple linear weighted combination of inputs. Instead, it effectively captures complex

nonlinear interactions and spatial dependencies between the Pangu and ECMWF forecasts.

Discussion

This study provides a systematic evaluation of boundary-layer jets (BLJs) forecasts over the northern South China Sea produced by a conventional numerical weather prediction (NWP) model (ECMWF) and a data-driven artificial intelligence (AI) model (Pangu-Weather). The assessments comprehensively span BLJ occurrence, overall low-level wind forecast skill, and key jet characteristics, including spatial extent, intensity, jet-core position, and wind direction. In addition, a deep-learning-based blending framework is developed to exploit the complementary error structures of the two models for improved prediction accuracy. The main conclusions are summarized as follows.

The AI-based Pangu model exhibits superior overall performance in forecasting the low-level wind field over the northern SCS, with consistently lower RMSE and higher spatial correlation coefficients than ECMWF. These results suggest that data-driven models trained on large reanalysis datasets can effectively capture the dominant statistical evolution of regional low-level winds.

Despite its higher overall accuracy, Pangu shows systematic biases that contrast sharply with those of ECMWF. ECMWF consistently overestimates BLJ intensity and spatial extent, whereas Pangu systematically underestimates BLJ intensity and scale. In addition, ECMWF forecasts show a northward displacement of the jet core and a clockwise wind-direction bias, while Pangu forecasts exhibit a northeastward shift accompanied by a counterclockwise bias. These opposing error patterns reveal fundamentally different limitations in physics-based and data-driven forecasting systems.

Pangu performs notably better in non-BLJ events, likely reflecting the tendency of data-driven models to regress toward the climatological mean and smooth extreme anomalies. In contrast, the physics-based ECMWF is better able to capture extreme wind events but suffers from pronounced diurnal oscillations in forecast errors, particularly during sunrise and sunset transitions. These error oscillations likely reflect limitations in boundary-layer parameterizations and their representation of rapid stability changes.

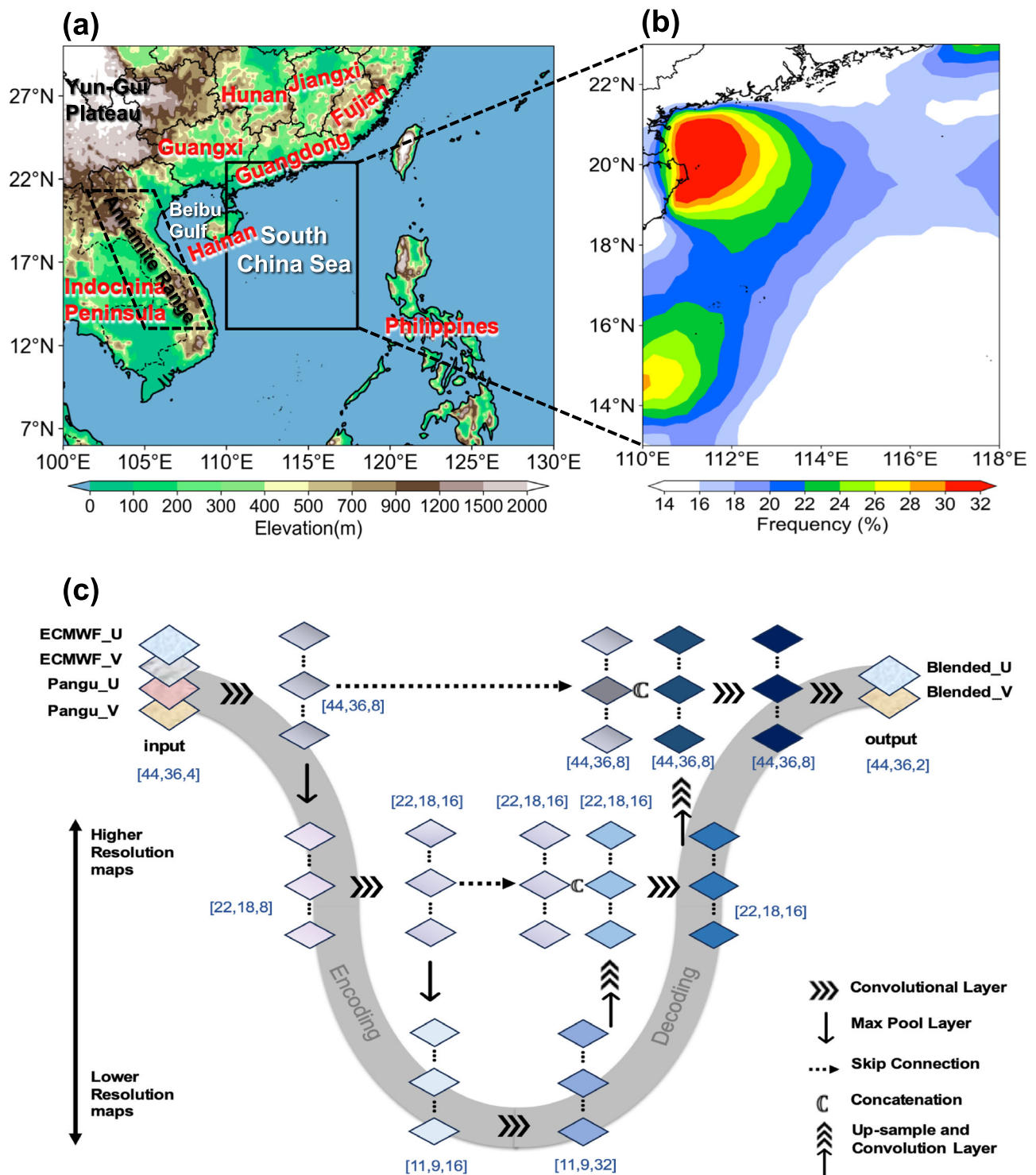


Fig. 6 | Study domain and U-Net-based blending framework. **a** Terrain height (shading, m) surrounding the SCS, with the black box indicating the study domain. **b** Climatological occurrence frequency of BLJs (shading, %) within the domain shown in (a). **c** The architecture of the U-Net-based blending model, adapted from the previous work⁴³. Zonal (U) and meridional (V) wind components from ECMWF

and Pangu are provided as inputs, and the network outputs blended U and V, optimized to approximate ERA5 wind distributions. Different arrow symbols represent different operations, as indicated in the legend. Numbers in blue parentheses indicate tensor dimensions in the form [x, y, channels].

By combining ECMWF and Pangu forecasts using a U-Net-based blending approach, forecast skill is significantly improved relative to either model alone. Importantly, this improvement is not solely due to the cancellation of complementary errors, as additional baseline experiments using linear regression and equal-weight averaging show that the U-Net consistently outperforms simple linear combinations. The success of the

blended forecasts arises from the ability of the deep-learning framework to extract multi-scale spatial features and from the implementation of a composite loss function that explicitly penalizes missed detections and false alarms of BLJ events. This loss function design effectively addresses the long-tailed distribution of wind speed, reducing overfitting to weak-wind conditions and improving robustness for high-intensity BLJ events. Crucially,

this targeted optimization does not degrade general forecasting skill, with the model also performing well under non-BLJ conditions. Importantly, this blended approach effectively mitigates the opposing systematic biases of the individual models, demonstrating that deep learning can serve as a powerful post-processing tool rather than a replacement for physics-based prediction.

While this study highlights robust and physically interpretable differences between AI-based and physics-based BLJ forecasts and demonstrates the value of exploiting their complementary errors, an important limitation remains. Pangu-Weather is used here as the sole representative of AI-based forecasting. Given the rapid development of the AI field, future work should extend the analysis to other emerging state-of-the-art models, such as Google DeepMind's GraphCast and ECMWF's own Artificial Intelligence/Integrated Forecasting System (AIFS). In addition, evaluating whether other high-resolution NWP models or ensembles exhibit similar systematic BLJ biases will be essential for establishing a more comprehensive benchmark. Such efforts will help clarify whether the complementary error structures identified here are model-specific or represent a more general feature of AI and physics-based approaches to BLJ forecasting. Future studies should also investigate the relative roles of diurnal and synoptic-scale forcing in shaping BLJ predictability across different modeling frameworks.

Methods

Data

The ECMWF forecast data used in this study are obtained from the THORPEX Interactive Grand Global Ensemble (TIGGE) archive. The dataset has a horizontal resolution of $0.25^\circ \times 0.25^\circ$ and provides forecasts twice daily at 00:00 and 12:00 UTC³⁵.

The AI-based forecasts are produced using the Pangu-Weather model, which operates at the same horizontal resolution of $0.25^\circ \times 0.25^\circ$. Pangu generates forecasts at 6-h intervals up to 4 days ahead using a combination of pre-trained 6-h and 24-h models to limit error growth with a minimum number of iterations. Importantly, Pangu is initialized from the same ECMWF operational analysis as the ECMWF forecasts, enabling a direct and consistent comparison between the two modeling systems.

Both forecasts are evaluated against the fifth-generation ECMWF global atmospheric reanalysis dataset (ERA5), which has been widely used and validated for the identification and analysis of coastal BLJs^{10,36}.

Definition of BLJs

Following widely adopted criteria^{4,37}, a BLJ is identified when (1) the maximum wind speed below 900 hPa exceeds 10 m s^{-1} , and (2) below 600 hPa, wind speed decreases by at least 3 m s^{-1} from the level of maximum wind to the minimum above it.

Based on the climatological high-frequency region of BLJs (Fig. 6a, b), a BLJ event over the northern SCS is defined as occurring when more than 18% of grid points within the domain outlined in Fig. 6a satisfy the BLJ criteria. Hourly ERA5 data are used to identify BLJ events at 00, 06, 12, and 18 UTC. For each identified event, ECMWF and Pangu forecasts are sampled at 6-h intervals from 6-h to 96-h lead times to assess predictive skill.

To characterize BLJ structure, four indices are defined to represent jet spatial extent, intensity, core location, and wind direction of the jet. Because Pangu provides only 13 vertical levels, with only two levels (925 hPa and 1000 hPa) available below 900 hPa, it cannot fully meet the vertical resolution requirements of the strict BLJ definition; we therefore adopt a wind speed threshold of 10 m s^{-1} at 925 hPa for all indices. The indices are defined as follows:

1. BLJ area index: The total area within $13^\circ\text{--}23^\circ\text{N}$ and $110^\circ\text{--}118^\circ\text{E}$ where the 925-hPa wind speed is $\geq 10 \text{ m s}^{-1}$.
2. BLJ intensity index: The mean 925-hPa wind speed of all grid points within the same region that exceed the 10 m s^{-1} threshold.
3. BLJ core position index: The latitude-longitude centroid of all 925-hPa grid points within the region that meet the threshold.
4. BLJ wind-direction index: The angle of the vector-mean wind derived from the meridional and zonal wind components over the same threshold-exceeding points.

Evaluation indicators

Wind speed forecast accuracy is evaluated using three widely employed metrics: the root-mean-square error (RMSE), the mean error (ME), and the standard deviation (Std). RMSE quantifies the overall magnitude of forecast errors²⁹, while ME diagnoses systematic biases. Std is introduced to assess the internal variability and dispersion of wind speed within each specific dataset. The formulas for these metrics are given as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^T \sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} L(i) (\hat{A}_{i,j,t} - A_{i,j,t})^2}{T \times N_{lat} \times N_{lon}}}, L(i) = N_{lat} \times \frac{\cos \phi_i}{\sum_{i=1}^{N_{lat}} \cos \phi_i} \quad (1)$$

$$ME = \frac{1}{T} \sum_{i=1}^T (\hat{A}_i - A_i) \quad (2)$$

$$Std = \sqrt{\frac{\sum_{t=1}^T \sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} (X_{i,j,t} - \bar{X})^2}{T \times N_{lat} \times N_{lon}}}, \bar{X} = \frac{\sum_{t=1}^T \sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} X_{i,j,t}}{T \times N_{lat} \times N_{lon}} \quad (3)$$

Here, $L(i)$ is the latitude-dependent weighting factor at latitude ϕ , $\hat{A}$ denotes the predicted wind speed, A represents the reference (observed) value, and T denotes the number of temporal samples. For the standard deviation, X represents the wind speed from a specific dataset (e.g., ERA5, ECMWF, or Pangu), while $\bar{X}$ denotes its corresponding spatiotemporal mean.

To further evaluate model performance in forecasting BLJs, categorical metrics are computed based on a wind speed threshold of 10 m s^{-1} , including the threat score (TS), miss rate (MR), and false alarm rate (FAR):

$$TS = \frac{TP}{TP + FP + FN} \quad (4)$$

$$MR = \frac{FN}{TP + FN} \quad (5)$$

$$FAR = \frac{FP}{TP + FP} \quad (6)$$

Here, TP denotes the number of grid points where both the forecast and the observation exceed the threshold, FN denotes the grid points where the observation exceeds the threshold but the forecast does not, and FP denotes the number of grid points where the forecast exceeds the threshold but the observation does not.

Blended forecasting architecture and loss function

To exploit the complementary error structures of ECMWF and Pangu forecasts, a deep-learning-based blending model is implemented using a U-Net architecture³⁸. The U-Net employs a symmetric encoder-decoder structure and has been widely applied in meteorological studies^{26,39,40} due to its capability to efficiently extract multi-scale spatial features via the encoder and reconstruct fine-scale structures through the decoder. Skip connections are utilized at each layer to concatenate encoder-derived high-level features with the decoder's low-level spatial details, thereby preserving spatial accuracy.

As depicted in Fig. 6c, the input tensor (44×36 grid points with 4 channels) is constructed from multi-source inputs and integrates complementary information from numerical weather forecasts (ECMWF) and AI-based predictions (Pangu-Weather) for both the zonal (U) and meridional (V) wind components. It then undergoes two stages of double convolution followed by downsampling. This process shrinks the feature map to 11×9 while expanding the number of feature channels to 32. The decoder mirrors this process through two upsampling and double convolution cycles, restoring the spatial resolution to 44×36 with 8 feature channels. A final 1×1 convolution layer produces a 2-channel output corresponding to the blended zonal and meridional wind components, without altering the spatial resolution.

To ensure an adequate sample size for model training, the BLJ event definition is relaxed to a 5% area threshold, and forecasts from 16 lead times are combined. This yields a total of 17,728 samples, including 14,856 training samples from April–June 2009–2017 and 2872 validation samples from April–June 2018–2019. An independent test period of April–June 2020–2024 is used to evaluate the blended model's performance. For consistency, the comparative evaluation between the Pangu-Weather and ECMWF forecasts is conducted over the same period.

To address limitations of conventional loss functions in BLJ forecasting, we design an improved composite loss function for the blended model. While a conventional U-Net trained with mean absolute error (MAE) effectively captures average wind conditions, MAE alone tends to bias predictions toward weak winds due to the long-tailed, non-Gaussian distribution of wind speed, resulting in underestimation of rare high-wind events. To mitigate this issue, we develop a composite loss function integrating MAE, MR, and FAR, explicitly penalizing missed and false BLJ detections. MR and FAR are computed from TP, FN, and FP, as defined earlier. However, because TP, FP, and FN are discrete quantities, they are inherently non-differentiable and cannot be directly optimized using a gradient-based method. To address this challenge, we employ a sigmoid function to approximate the step function. This transforms the discrete categorical metrics into continuous and differentiable formulations (TP_{diff} , FN_{diff} , and FP_{diff}), enabling MR and FAR to be seamlessly incorporated into gradient-based optimization during network training^{41,42}.

$$TP_{diff} = \frac{1}{N} \sum_{i=1}^N \left[(y_{obs} > \alpha) \odot \text{Sigmoid}(y_{pred} - \alpha) \right] \quad (7)$$

$$FN_{diff} = \frac{1}{N} \sum_{i=1}^N \left[(y_{obs} > \alpha) \odot \text{Sigmoid}(\alpha - y_{pred}) \right] \quad (8)$$

$$FP_{diff} = \frac{1}{N} \sum_{i=1}^N \left[(y_{obs} < \alpha) \odot \text{Sigmoid}(y_{pred} - \alpha) \right] \quad (9)$$

$$Loss = MAE + \lambda MR_{diff} + \mu FAR_{diff} \quad (10)$$

where $\odot$ denotes element-wise multiplication; α represents the wind speed threshold, which is set to 10 m s^{-1} in this study; and λ and μ are hyperparameters determined empirically. Based on a comprehensive evaluation of model performance on the validation set, the values $\lambda = 2.11$ and $\mu = 0.48$ are identified as optimal, as they jointly minimize the RMSE and the errors of key BLJ-related indices.

To quantitatively evaluate the relative contributions of individual forecast sources (ECMWF and Pangu-Weather) and their respective spatial components to the blended output, we employ the SHapley Additive exPlanations (SHAP) framework. SHAP is a model-agnostic methodology rooted in cooperative game theory that interprets complex machine learning models by assigning an importance value to each input feature. In this context, the blending process is conceptualized as a “game”, where each explanatory variable (e.g., the zonal and meridional wind components from different models) acts as a “player” and the final blended wind field represents the “payout”. By calculating the marginal contribution of each feature across all possible feature combinations, SHAP provides a quantitative measure of its respective influence.

In parallel with the development of the U-Net framework, two baseline blending methods are implemented to determine whether the proposed deep-learning approach provides substantial benefits beyond simple statistical techniques. First, an equally weighted average (Equal-Weight) of the ECMWF and Pangu forecasts is computed, representing a simple error-cancellation strategy. Second, a linear regression (Linear) model is trained to optimally combine the two forecasts by learning linear weights that minimize the prediction error with respect to the reference data. The linear regression blending model is trained using the same training dataset as the

U-Net-based model. These baseline methods provide a benchmark to distinguish improvements arising from the U-Net learning capability from those due to simple averaging of forecasts with opposing systematic biases.

Data availability

The ERA5 reanalysis data are obtained from the Copernicus Data Store (<https://cds.climate.copernicus.eu/>). The ECMWF operational forecast data are accessed through TIGGE (<https://apps.ecmwf.int/datasets/data/tigge/levtype=pl/type=cf/>). The Pangu-Weather model is available via its official GitHub repository (<https://github.com/198808xc/Pangu-Weather>). The Blended forecast data used for the figures and analysis are available at <https://doi.org/10.5281/zenodo.18606531>.

Code availability

The codes that support the findings of this study are available from the corresponding author on request.

Received: 14 February 2026; Accepted: 12 April 2026;

Published online: 22 April 2026

References

1. Stensrud, D. J. Importance of low-level jets to climate: a review. *J. Clim.* **9**, 1698–1711 (1996).
2. Du, Y. & Chen, G. Heavy rainfall associated with double low-level jets over southern China. Part I: ensemble-based analysis. *Mon. Weather Rev.* **146**, 3827–3844 (2018).
3. Du, Y. & Chen, G. Heavy rainfall associated with double low-level jets over southern China. Part II: convection initiation. *Mon. Weather Rev.* **147**, 543–565 (2019).
4. Du, Y. & Rotunno, R. A simple analytical model of the nocturnal low-level jet over the Great Plains of the United States. *J. Atmos. Sci.* **71**, 3674–3683 (2014).
5. Rife, D. L., Pinto, J. O., Monaghan, A. J., Davis, C. A. & Hannan, J. R. Global distribution and characteristics of diurnally varying low-level jets. *J. Clim.* **23**, 5041–5064 (2010).
6. Blackadar, A. K. Boundary layer wind maxima and their significance for the growth of nocturnal inversions. *Bull. Am. Meteorol. Soc.* **38**, 283–290 (1957).
7. Helfand, H. M. & Schubert, S. D. Climatology of the simulated Great Plains low-level jet and its contribution to the continental moisture budget of the United States. *J. Clim.* **8**, 784–806 (1995).
8. Holton, J. R. The diurnal boundary layer wind oscillation above sloping terrain. *Tellus* **19**, 199–205 (1967).
9. Du, Y., Chen, Y.-L. & Zhang, Q. Numerical simulations of the boundary layer jet off the southeastern coast of China. *Mon. Weather Rev.* **143**, 1212–1231 (2015).
10. Du, Y., Shen, Y. & Chen, G. Influence of coastal marine boundary layer jets on rainfall in South China. *Adv. Atmos. Sci.* **39**, 782–801 (2022).
11. Huang, X. et al. Uplift mechanism of coastal extremely persistent heavy rainfall (EPHR): the key role of low-level jets and ageostrophic winds in the boundary layer. *Geophys. Res. Lett.* **49**, e2021GL096029 (2022).
12. Zhang, M. & Meng, Z. Impact of synoptic-scale factors on rainfall forecast in different stages of a persistent heavy rainfall event in South China. *J. Geophys. Res. Atmos.* **123**, 3574–3593 (2018).
13. Du, Y. & Chen, G. Climatology of low-level jets and their impact on rainfall over southern China during the early-summer rainy season. *J. Clim.* **32**, 8813–8833 (2019).
14. Li, X. & Du, Y. Statistical relationships between two types of heavy rainfall and low-level jets in South China. *J. Clim.* **34**, 8549–8566 (2021).
15. Shen, Y. & Du, Y. Sensitivity of boundary layer parameterization schemes in a marine boundary layer jet and associated precipitation during a coastal warm-sector heavy rainfall event. *Front. Earth Sci.* **10**, 1085136 (2023).
16. Chen, G., Sha, W., Iwasaki, T. & Wen, Z. Diurnal cycle of a heavy rainfall corridor over East Asia. *Mon. Weather Rev.* **145**, 3365–3389 (2017).

17. Trier, S. B. & Parsons, D. B. Evolution of environmental conditions preceding the development of a nocturnal mesoscale convective complex. *Mon. Weather Rev.* **121**, 1078–1098 (1993).
18. Dong, F., Zhi, X., Zhang, L. & Ye, C. Diurnal variations of coastal boundary layer jets over the northern South China Sea and their impacts on diurnal cycle of rainfall over southern China during the early-summer rainy season. *Mon. Weather Rev.* **149**, 3341–3363 (2021).
19. Du, Y. et al. Convection initiation and growth at the coast of South China. Part I: effect of the marine boundary layer jet. *Mon. Weather Rev.* **148**, 3847–3869 (2020).
20. Zhang, M. & Meng, Z. Warm-sector heavy rainfall in southern China and its WRF simulation evaluation: a low-level-jet perspective. *Mon. Weather Rev.* **147**, 4461–4480 (2019).
21. Lin, Y. & Du, Y. Sensitivity of coastal rainfall simulation to the assimilation of wind profiles near the marine boundary layer jet: an observing system simulation experiment. *J. Geophys. Res. Atmos.* **130**, e2025JD044106 (2025).
22. Bauer, P., Thorpe, A. & Brunet, G. The quiet revolution of numerical weather prediction. *Nature* **525**, 47–55 (2015).
23. Chen, L. et al. FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. *npj Clim. Atmos. Sci.* **6**, 190 (2023).
24. Dueben, P. D. & Bauer, P. Challenges and design choices for global weather and climate models based on machine learning. *Geosci. Model Dev.* **11**, 3999–4009 (2018).
25. Lam, R. et al. Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).
26. Weyn, J. A., Durran, D. R. & Caruana, R. Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *J. Adv. Model. Earth Syst.* **12**, e2020MS002109 (2020).
27. Pathak, J. et al. FourCastNet: a global data-driven high-resolution weather model using adaptive Fourier neural operators. <https://arxiv.org/abs/2202.11214> (2022).
28. Zhang, Y. et al. Skilful nowcasting of extreme precipitation with NowcastNet. *Nature* **619**, 526–532 (2023).
29. Bi, K. et al. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**, 533–538 (2023).
30. Sandu, I., Bechtold, P., Nuijens, L., Beljaars, A. & Brown, A. On the causes of systematic forecast biases in near-surface wind direction over the oceans. *ECMWF Tech. Memo.* **866**, 1–21 (2020).
31. Heppelmann, T., Steiner, A. & Vogt, S. Application of numerical weather prediction in wind power forecasting: Assessment of the diurnal cycle. *Meteorol. Z.* **26**, 319–331 (2017).
32. Holtzlag, A. A. M. et al. Stable atmospheric boundary layers and diurnal cycles: challenges for weather and climate models. *Bull. Am. Meteorol. Soc.* **94**, 1691–1706 (2013).
33. Mirocha, J. D., Simpson, M. D., Fast, J. D., Berg, L. K. & Baskett, R. L. Investigation of boundary-layer wind predictions during nocturnal low-level jet events using the weather research and forecasting model: Evaluating mesoscale low-level jet simulations using the WRF model. *Wind Energy* **19**, 739–762 (2016).
34. Sandu, I., Beljaars, A. & Balsamo, G. Improving the representation of stable boundary layers. *ECMWF Newsl.* **138**, 24–29 (2014).
35. Zeng, Z. et al. Enhancing short-term wind speed prediction capability of numerical weather prediction through machine learning methods. *J. Geophys. Res. Atmos.* **129**, e2024JD041822 (2024).
36. Kalverla, P. C., Duncan, J. B. Jr., Steeneveld, G.-J. & Holtzlag, A. A. M. Low-level jets over the North Sea based on ERA5 and observations: together they do better. *Wind Energy Sci.* **4**, 193–209 (2019).
37. Tu, C.-C., Chen, Y.-L., Lin, P.-L. & Du, Y. Characteristics of the marine boundary layer jet over the South China Sea during the early summer rainy season of Taiwan. *Mon. Weather Rev.* **147**, 457–475 (2019).
38. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. *Med. Image Comput. Comput. Assist. Interv.* **9351**, 234–241 (2015).
39. Hess, P. & Boers, N. Deep learning for improving numerical weather prediction of heavy rainfall. *J. Adv. Model. Earth Syst.* **14**, e2021MS002765 (2022).
40. Hu, Y., Yin, F. & Zhang, W. Deep learning-based precipitation bias correction approach for Yin–He global spectral model. *Meteorol. Appl.* **28**, e2032 (2021).
41. Larraondo, P. R., Renzullo, L. J., Van Dijk, A. I. J. M., Inza, I. & Lozano, J. A. Optimization of deep learning precipitation models using categorical binary metrics. *J. Adv. Model. Earth Syst.* **12**, e2019MS001909 (2020).
42. You, X., Liang, Z., Wang, Y. & Zhang, H. A study on loss function against data imbalance in deep learning correction of precipitation forecasts. *Atmos. Res.* **281**, 106500 (2023).
43. Chase, R. J., Harrison, D. R., Lackmann, G. M. & McGovern, A. A machine learning tutorial for operational meteorology. Part II: neural networks and deep learning. *Weather Forecast* **38**, 1271–1293 (2023).

Acknowledgements

This study was supported by National Key Research and Development Program of China (Grant 2024YFC3013003), the Heavy Rainfall Research Foundation of China (BYKJ2024Z01), the National Natural Science Foundation of China (Nos. 42475002 and 42122033), the Guangdong Major Project of Basic and Applied Basic Research (2024A1515510005, 2025A1515011974, and 2020B0301030004), the Project supported by Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. SML2024SP035, SML2024SP012, and 311024001), the Key Innovation Team of China Meteorological Administration (CMA2023ZD08). We also acknowledge the high-performance computing support from the School of Atmospheric Sciences of Sun Yat-sen University.

Author contributions

H.L. wrote the manuscript, developed the deep-learning model, and conducted the data processing. Y.D. conceived the research, supervised the project, and contributed to the discussion and manuscript revision. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Yu Du.

Reprints and permissions information is available at

<http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026