ADVANCES IN ATMOSPHERIC SCIENCES, 2026

• Original Paper •

A Synergistic Approach: Dynamics–AI Ensemble in Tropical Cyclone Forecasting

Yonghui LI[1,2,3], Wansuo DUAN*[1,2], Hao LI*[4,3], Wei HAN[5,6], Han ZHANG[7], and Yinuo LI[8]

[1]*State Key Laboratory of Earth System Numerical Modeling and Application*, *Institute of Atmospheric Physics*, *Chinese Academy of Sciences*, *Beijing* 100029, *China*

[2]*University of Chinese Academy of Sciences*, *Beijing* 101408, *China*

[3]*Shanghai Academy of Artificial Intelligence for Science (SAIS)*, *Shanghai* 200232, *China*

[4]*Artificial Intelligence Innovation and Incubation Institute*, *Fudan University*, *Shanghai* 201203, *China*

[5]*State Key Laboratory of Severe Weather Meteorological Science and Technology (LASW)*, *Beijing* 100081, *China*

[6]*CMA Earth System Modeling and Prediction Centre (CEMC)*, *China Meteorological Administration*, *Beijing* 100081, *China*

[7]*Henan Meteorological Observatory*, *Zhengzhou* 450003, *China*

[8]*Tianjin Meteorological Information Center*, *Tianjin* 300074, *China*

(Received 11 November 2025; revised 16 March 2026; accepted 24 April 2026)

ABSTRACT

This study addresses a critical challenge in AI-based weather forecasting by developing an AI-driven optimized ensemble forecast system using Orthogonal Conditional Nonlinear Optimal Perturbations (O-CNOPs). The system bridges the gap between computational efficiency and dynamic consistency in tropical cyclone (TC) forecasting. Unlike conventional ensembles limited by computational costs or AI ensembles constrained by inadequate perturbation methods, O-CNOPs generate dynamically optimized perturbations that capture fast-growing errors of the FuXi model while maintaining plausibility. The key innovation lies in producing orthogonal perturbations that respect FuXi's nonlinear dynamics, yielding structures reflecting dominant dynamical controls and physically interpretable probabilistic forecasts. Demonstrating superior deterministic and probabilistic skills over the operational Integrated Forecasting System Ensemble Prediction System, this work establishes a new paradigm combining AI's computational advantages with rigorous dynamical constraints. Success in TC track forecasting paves the way for reliable ensemble forecasts of other high-impact weather systems, marking a major step toward operational AI-based ensemble forecasting.

Key words: tropical cyclone, initial uncertainty, ensemble forecasting, artificial intelligence

Citation: Li, Y. H., W. S. Duan, H. Li, W. Han, H. Zhang, and Y. N. Li, 2026: A synergistic approach: Dynamics–AI ensemble in tropical cyclone forecasting. *Adv. Atmos. Sci.*, https://doi.org/10.1007/s00376-026-5784-3.

Article Highlights:

- An AI-driven ensemble system is developed using O-CNOPs to generate physically consistent, optimized perturbations for TC forecasts.
- O-CNOPs capture realistic forecast uncertainties for TC tracks in the FuXi model, enabling reliable forecasts.
- The system generally outperforms operational forecasts in deterministic and probabilistic scores.

1. Introduction

Tropical cyclones (TCs) are often accompanied by extreme weather events such as strong winds, torrential rains, massive waves, and storm surges, causing severe damage to life and property (Peduzzi et al., 2012; Zhang et al., 2017; Gori et al., 2022). It is therefore extremely important for disaster prevention and reduction to forecast TCs accurately. However, due to the highly chaotic nature of TC systems (Wang and Wu, 2004; Nystrom et al., 2018), even minor errors in initial conditions can rapidly amplify over time, leading to significant forecast uncertainty (Heming, 2017; Zhang et al., 2023; Li et al., 2025), as found in the Lorenz system (Lorenz, 1963, 1969). To address this, forecasters typically use the ensemble forecasting strategy, which generates multiple possible scenarios to quantify this uncertainty, thereby improving the forecast reliability and provid-

* Corresponding authors: Wansuo DUAN, Hao LI
Emails: duanws@lasg.iap.ac.cn, lihao_lh@fudan.edu.cn

© Institute of Atmospheric Physics/Chinese Academy of Sciences, and Science Press 2026

ing a scientific basis for disaster prevention decision-making.

The ensemble strategy has substantially enhanced TC forecasting performance in major meteorological departments such as ECWMF, NECP, etc. (Mureau et al., 1993; Toth and Kalnay, 1993; Bourke et al., 1995; Houtekamer et al., 1996), but it demands extensive computational resources due to reliance on complex physics-based numerical models. Nevertheless, it is very exciting that high-efficiency artificial intelligence (AI) techniques have emerged in recent years as powerful tools for forecasting-model development. Several AI-driven large meteorological models—including FuXi (Chen et al., 2023), Pangu-Weather (Bi et al., 2023), AIFS (Lang et al., 2024), GraphCast (Lam et al., 2023) etc.—have attracted considerable attention for their exceptional computational efficiency and predictive accuracy. These models present viable alternatives to traditional numerical approaches. Particularly, Li et al. (2025) adopted the FuXi model to explore TC track predictability and found that initial perturbations exhibit a significant growth phenomenon; furthermore, Pu et al. (2025) revealed that initial perturbations of particular spatial structure enhance such error growth and are favorable for ensemble forecasts. That is to say, similar to the Lorenz system (Lorenz, 1963, 1969), the output of the FuXi model exhibits extreme sensitivity to initial input data uncertainties. This discovery unlocks a transformative opportunity: by integrating dynamical ensemble methods with AI architectures like FuXi, one can achieve quantum leaps in both computational efficiency and predictive reliability, fundamentally overcoming the trade-offs of traditional approaches.

Currently, it is popular for ensemble forecasts that incorporate rapidly growing initial perturbations with physics-informed models to generate ensemble members. The major methods to yield rapidly growing initial perturbations are singular vectors (Mureau et al., 1993; Buizza and Palmer, 1995) (SVs) and breeding vectors (Toth and Kalnay, 1993) (BVs); however, these two methods either rely on linear dynamics and cannot depict the nonlinear nature of atmospheric and oceanic motions, or are uncertain in their representation of fast-growing perturbations during the forecast period, which indeed hinders the enhancement of ensemble forecast capability (Duan et al., 2023; Zhang et al., 2023). Regarding this limitation, a novel method of orthogonal conditional nonlinear optimal perturbations (Duan and Huo, 2016) (O-CNOPs) has been proposed to generate the fast-growing initial perturbations for ensemble forecasts. This method fully accounts for nonlinear physical processes and has been shown to possess better performance than SVs and BVs in TC track forecasting (Huo and Duan, 2019; Zhang et al., 2023). With respect to the FuXi model, it is a point of concern as to whether its integration with O-CNOPs can also significantly enhance the performance in TC track ensemble forecasting, besides its expected high efficiency.

In this study, we integrated O-CNOPs with the FuXi model and established a new ensemble forecasting system (hereafter referred to as "FuXi_CNOP") to forecast TC tracks. Following the operational practice of system-to-system benchmarking (e.g., NCEP vs. ECMWF), we compared its performance against world-leading systems. We reveal that, with only a small number of members produced, it can achieve markedly improved forecasting capability with respect to TC tracks—even higher than the world-leading ensemble forecast system. Accordingly, this study provides a highly promising approach to produce ensembles for AI model forecasts of TC tracks.

2. Methods and experimental setup

2.1. *O-CNOPs*

O-CNOPs (Duan and Huo, 2016) are a group of mutually orthogonal initial perturbations with maximum growth in their respective subspaces:

$$J\left(\delta \mathbf{x}_{0j}^{*}\right) = \max_{\delta \mathbf{x}_{0j} \in \Omega_j} \| \mathrm{M}_\tau\left(\mathbf{x}_0 + \delta \mathbf{x}_{0j}\right) - \mathrm{M}_\tau(\mathbf{x}_0) \|_{\mathrm{C}_2}, \quad (1)$$

where

$$\Omega_j = \begin{cases} \delta \mathbf{x}_{0j}^{*} \in \mathbb{R}^n \ ||| \ \delta \mathbf{x_0} \|_{\mathrm{C}_1} \leqslant \varepsilon, j = 1 \\ \delta \mathbf{x}_{0j}^{*} \in \mathbb{R}^n \ ||| \ \delta \mathbf{x_0}_j \|_{\mathrm{C}_1} \leqslant \varepsilon, \delta \mathbf{x}_{0j} \perp \Omega_k, k = 1, \cdots, j-1, j > 1 \end{cases}. \quad (2)$$

Here, M_τ is a nonlinear operator that evolves a state from initial time $t = 0$ to a future forecast time $t = \tau$; $\mathbf{x}_0$ is an initial condition and $\delta \mathbf{x}_0$ is its initial perturbation; $\mathbb{R}^n$ denotes n-dimensional Euclidean space; Ω_j is one of subspaces in $\mathbb{R}^n$ and defines an initial perturbation space; the norm $\| \cdot \|_{\mathrm{C}_1}$ is the measure of the initial perturbation; ε is the threshold for the amplitude of the perturbation and $\| \cdot \|_{\mathrm{C}_2}$ is the measure of its evolution at the forecast time; and the symbol "$\perp$" signifies the orthogonality of vector spaces. According to Eqs. (1) and (2), the 1st CNOP has the maximum nonlinear growth in the first subspace, i.e., the entire perturbation phase space Ω_1; and the jth CNOP provides the maximum nonlinear growth in the subspace orthogonal to the first $j-1$ CNOPs. O-CNOPs can be seen as an extension of traditional SVs, employed in the IFS_EPS in ECMWF, in the nonlinear regime, fully taking into account the impact of nonlinear physical processes.

In the present study, the initial perturbation $\delta \mathbf{x}_0$ is composed of $\mathbf{u}_0^{'}$, $\mathbf{v}_0^{'}$, $\mathbf{T}_0^{'}$, which correspond to zonal and meridional winds and temperature, respectively. The norm $\| \cdot \|_{\mathrm{C}_1}$ is taken as the sum of kinetic and internal energy [see Eq. (3)] and measures the amplitude of initial perturbations, while $\| \cdot \|_{\mathrm{C}_2}$ measures the perturbation kinetic energy at the forecast time τ [see Eq. (4)]. Specifically, they have the expressions as follows:

$$\| \delta \mathbf{x}_0 \|_{\mathrm{C}_1}^2 = \frac{1}{D_1} \int_\sigma \int_{D_1} \left[\mathbf{u}_0^{\prime 2} + \mathbf{v}_0^{\prime 2} + \frac{c_p}{T_r} \mathbf{T}_0^{\prime 2} \right] \mathrm{d}D_1 \mathrm{d}\sigma, \quad (3)$$

and

$$
\begin{aligned}
J &= \| \mathrm{M}_\tau(\mathbf{x}_0 + \delta\mathbf{x}_0) - \mathrm{M}_\tau(\mathbf{x}_0) \|_{C_2} \\
&= \frac{1}{D_2} \int_\sigma \int_{D_2} \left[\mathbf{u}_\tau'^2 + \mathbf{v}_\tau'^2 \right] \mathrm{d}D_2 \mathrm{d}\sigma \,,
\end{aligned}
\tag{4}
$$

where c_p is the specific heat at a constant pressure, with a value of 1005.7 J kg^{-1} K^{-1}; the reference parameter $T_r = 270$ K; $\mathbf{u}_0'$ and $\mathbf{v}_0'$ are the wind components of the initial perturbation; $\mathbf{T}_0'$ is the temperature component; D_1 covers the global atmosphere and σ is the vertical coordinate from 13 pressure levels of the FuXi model; J is the objective function defined as the kinetic energy over the verification area D_2, which covers the corresponding TC's initial position and entire track trajectory with uniform 5° expansion in all cardinal directions, in an attempt to fully interpret the positional uncertainties of the TC circulation (Zhang et al., 2023); and $\mathbf{u}_\tau'$ and $\mathbf{v}_\tau'$ are the evolution of the wind components of the initial perturbation $\delta\mathbf{x}_0$ at the forecast time τ.

Applying Eqs. (3) and (4) to Eq. (1), one can calculate the first CNOP through the automatic differentiation module that robustly exists in AI models—a process that implements the objective function using PyTorch's computational graph (Paszke et al., 2017) and utilizes the automatic differentiation module to calculate the gradient of the objective function (Li et al., 2025) (see Fig. 1a). Then, in the subspace orthogonal to the first CNOP $\delta\mathbf{x}_{01}^*$, one computes the second CNOP $\delta\mathbf{x}_{02}^*$; in the subspace orthogonal to both the first and second CNOPs, one computes the third CNOP $\delta\mathbf{x}_{03}^*$, ..., and in the subspace orthogonal to the first $j-1$ CNOPs, one can calculate the jth CNOP $\delta\mathbf{x}_{0j}^*$ (see Fig. 1b). Then, a group of O-CNOPs can be achieved for ensemble forecasts.

2.2. *FuXi model*

The FuXi model (Chen et al., 2023) was developed by the Artificial Intelligence Innovation and Incubation Institute at Fudan University. It employs a cascaded neural architecture to generate 15-day global atmosphere forecasts at 0.25° spatial and 6-hour temporal resolutions. It incorporates 70 state variables, utilizing input data combining both the current and previous time steps with dimensions of 2×70×721×1440. This model is trained utilizing a 36-year (1979–2015) ERA5 reanalysis dataset. Chen et al. (2023) demonstrated that the forecast performance, evaluated using latitude-weighted RMSEs and anomaly correlation coefficients, is comparable to the ECMWF ensemble mean even at a 15-day lead time, making it the first AI-based model to achieve such accuracy.

2.3. *Experimental strategy*

The ensemble forecasting skill is sensitive to ensemble parameters including the amplitudes of initial perturbations, the optimization time interval, and the ensemble size N (Toth and Kalnay, 1997; Duan et al., 2023). Based on strategic sensitivity experiments [see Fig. S1 in the electronic supplementary material (ESM); also refer to Zhang et al. (2023)], the ensemble parameters for FuXi_CNOP were determined as perturbation amplitude $\varepsilon = 0.21$ J kg^{-1}, optimization time interval $\tau = 0.5$ days, and ensemble size 31, which comprise one control forecast and 30 ensemble members produced by superimposing 15 pairs of positive and negative CNOPs onto the control forecast. The control forecasts are generated by the FuXi model, which are initialized by 12-hour lead-time forecast fields starting from ERA5 reanalysis data.

The study evaluates TC track ensemble forecasts gener-

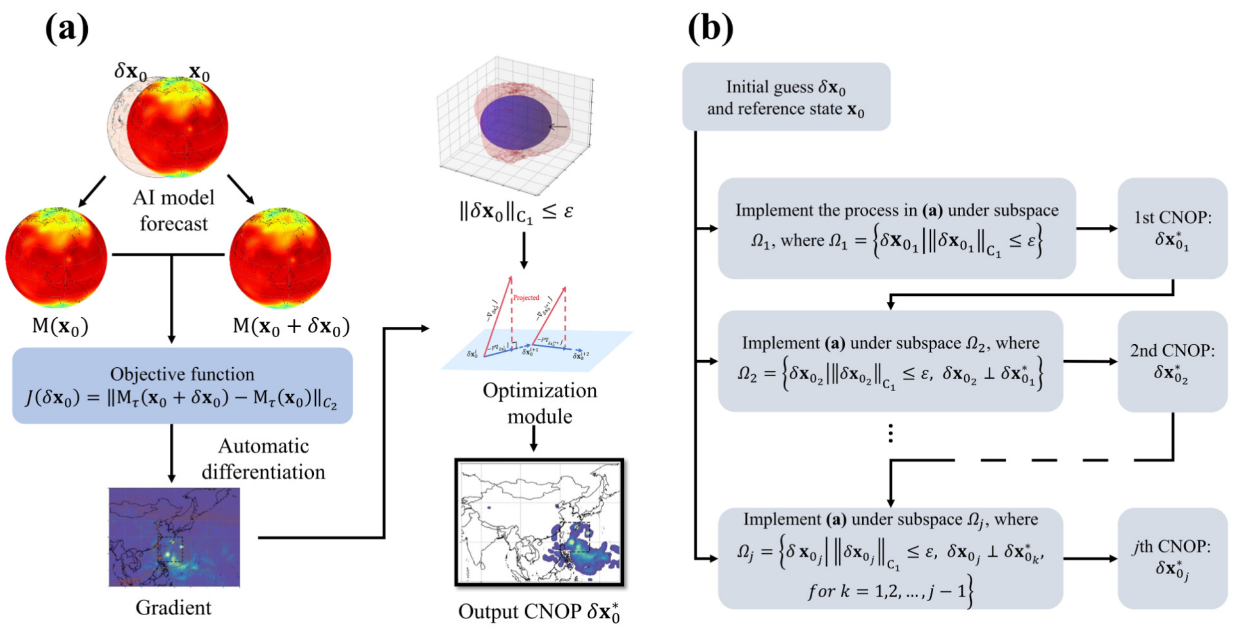

Fig. 1. The (a) process of searching for CNOPs using the automatic differentiation module in an AI model, and (b) computational flow of O-CNOPs. ε is the threshold for the amplitude of the perturbation.

ated by FuXi_CNOP as compared with the Integrated Forecast System Ensemble Prediction System (IFS_EPS) of ECMWF—the current benchmark for numerical weather prediction–based ensemble forecasting (Conroy et al., 2023). We utilize the TIGGE dataset containing ECMWF's IFS version 48r1 forecasts (native 9 km/137 levels, interpolated to 0.25° resolution). The TC best-track dataset used is the IBTrACS dataset. A comprehensive description of the TC tracker can be found in Zhong et al. (2024a). This comparative framework enables rigorous assessment of both deterministic and probabilistic forecast skills while accounting for realistic initialization scenarios.

We conducted comprehensive diagnostic comparisons between FuXi_CNOP and IFS_EPS. The analysis focuses on two primary TC basins: the western North Pacific and North Atlantic. To ensure rigorous out-of-sample validation, our analysis focuses exclusively on TC events occurring during 2018–23, which is a period completely independent from FuXi's training period of 1979–2015. The study assesses TC track forecast skill through 91 forecasts of 62 TCs, including 51 cases of 35 TCs over the western North Pacific and 40 cases of 27 TCs over the North Atlantic, all of whose intensities reached the tropical storm criterion. This evaluation framework enables direct comparison between FuXi_CNOP and IFS_EPS under identical verification conditions.

3. Results

3.1. *Comparative analysis of FuXi_CNOP and IFS_EPS in TC deterministic forecasting*

In this section, we perform a comprehensive comparison between IFS_EPS and our newly developed FuXi_CNOP, as based on an analysis of a total of 91 cases, including 51 cases occurring in the Northwest Pacific and 40 cases arising in the North Atlantic. Note that IFS_EPS employs SVs and ensemble data assimilation to generate initial perturbations, which, combined with stochastic physics perturbations, account for both the initial and model uncertainties of control forecasts generated by the IFS, with 51 ensemble members. However, our FuXi_CNOP only generates initial perturbations using the O-CNOPs method and was specifically designed to estimate initial uncertainties through 31 ensemble numbers.

Figure 2 presents a comparative evaluation of FuXi_CNOP and IFS_EPS in TC track forecasting, focusing on ensemble mean forecast errors, ensemble spread, along-track errors (ATE), and cross-track errors (CTE) (see Appendix A). In the short lead-time (0–12 h) forecasts, both systems exhibit comparable performance. However, beyond 18 h, FuXi_CNOP demonstrates a significant reduction in forecast errors, significantly outperforming IFS_EPS from 24 h onward and maintaining this advantage up to 120 h. This improvement is statistically significant ($p < 0.1$) between 36 and 120 h, with a maximum error reduction of 32.33%. Notably, FuXi_CNOP achieves a near-unity skill-spread ratio, mirroring the reliability of IFS_EPS in characterizing forecast uncertainty. This dual capability—enhanced accuracy and robust uncertainty quantification—highlights its superiority in long-range TC track forecasts.

In addition, FuXi_CNOP demonstrates substantially lower ATE and CTE compared to IFS_EPS beyond the 24-hour lead time. In particular, FuXi_CNOP maintains near-zero ATE values, reflecting precise translation speed estimation, whereas IFS_EPS exhibits persistent negative error, indicative of systematic underestimation of TC motion velocity. In terms of CTE performance, FuXi_CNOP outperforms IFS_EPS, with an indication of significantly reduced leftward track biases. The scatter distributions of ATE–CTE phase points for all 91 cases (see Figs. 2d, e) reveal two distinct patterns: FuXi_CNOP's phase points cluster tightly around the coordinate origin, with the one averaged for 91 cases positioned close to the coordinate origin, confirming trivially unbiased performance in TC motion; while in stark contrast, IFS_EPS phase points mostly concentrate in the third quadrant, with the one averaged for all TCs displaced accordingly—a spatial signature of combined slow-speed bias and consistent leftward track deviations in predicted TC tracks.

It is obvious that, for the deterministic forecasts in terms of the ensemble mean, ensemble spread, ATE and CTE, FuXi_CNOP aggressively possesses more advantages than IFS_EPS in not only forecasting accuracy but also performance consistency for TC tracks. We further categorized TCs into two distinct basins, i.e., the above-mentioned western North Pacific (51 cases) and North Atlantic (40 cases), with consistent results observed (see Figs. S2 and S3 in the ESM).

3.2. *Comparative analysis on uncertainty quantification provided by FuXi_CNOP and IFS_EPS in TC track probability forecasting*

The continuous ranked probability score (CRPS) is a rigorously defined statistical metric for quantifying the accuracy of probabilistic forecasts by measuring the disparity between a predicted cumulative distribution function and the observed value of a continuous variable (see Appendix A). Lower CRPS values denote superior performance. In this study, we apply the CRPS to assess the probabilistic forecasting skill of TC track predictions by comparing model outputs against best-track observations. Figure 3 presents comparative CRPS results for FuXi_CNOP and IFS_EPS across forecast lead times. It shows that, for the lead times of 0–12 hours, the two systems demonstrate comparable CRPS values; while beyond 18 hours, FuXi_CNOP shows obvious advantages, especially from 24 hours, overtaking IFS_EPS, and through 120 hours, maintaining consistent superiority (see Fig. 3a), achieving a maximum CRPS reduction of 29.2% (see Fig. 3b). These advantages for FuXi_CNOP are statistically significant ($p < 0.1$) for 36–120-hour lead times, with most lead times reaching $p < 0.05$. These findings demonstrate that FuXi_CNOP not only improves the accuracy of ensemble-mean track forecasts but also provides a reliable representation of forecast uncertainty (i.e., the first and

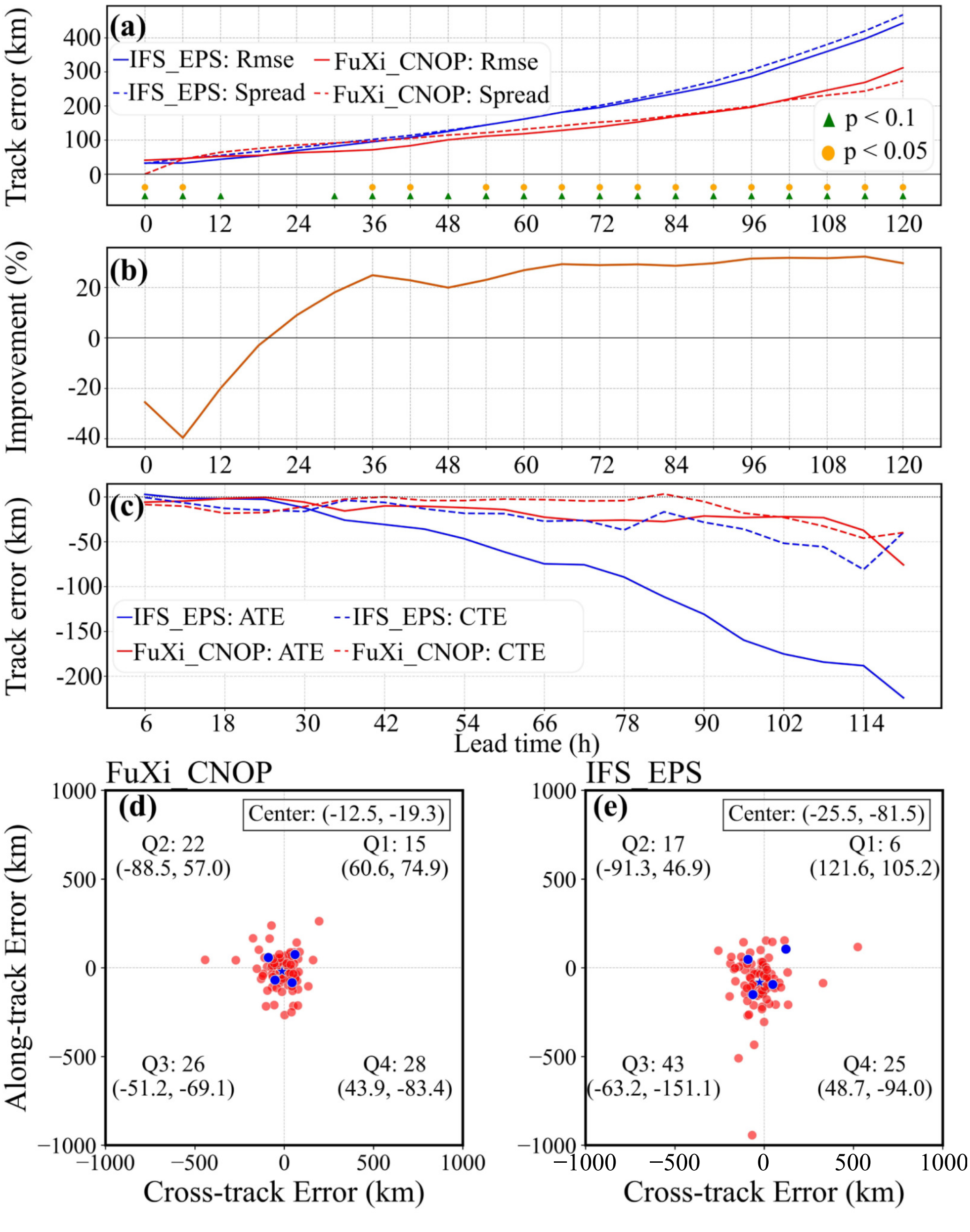

Fig. 2. Evaluation of deterministic forecasting skills. (a) Ensemble mean track errors (solid lines) and spread (dashed lines) for FuXi_CNOP (red) and IFS_EPS (blue) based on 91 cases. Yellow circles and green triangles indicate statistically significant differences at 95% and 90% confidence levels, respectively (Student's *t*-test). (b) Percentage reduction in ensemble mean track forecast errors of FuXi_CNOP relative to IFS_EPS. (c) Mean along-track errors (ATE, solid lines) and cross-track errors (CTE, dashed lines) as a function of lead time, averaged across 91 cases. (d) Scatter distribution of ATE vs. CTE for FuXi_CNOP, with quadrant means (blue dots) and overall means (blue stars) shown for reference. (e) As in (d) but for IFS_EPS.

second moments of the probability density distribution), thereby delivering a more skillful probabilistic characterization of TC track evolution at extended lead times. We also categorized TCs into western North Pacific and North Atlantic cases, and consistent results were obtained (see Figs. S4 and S5 in the ESM).

More specifically, we explored the TC track strike probability in ensemble forecasts generated by FuXi_CNOP and IFS_EPS, where both the Brier Score (BS) and Receiver Operating Characteristic (ROC) curve were adopted to measure the probability. We found that FuXi_CNOP achieves BS values comparable to those of IFS_EPS (see Fig. S6 in the ESM), while its ROCA (receiver operating characteristic area) performance is marginally inferior to that of IFS_EPS

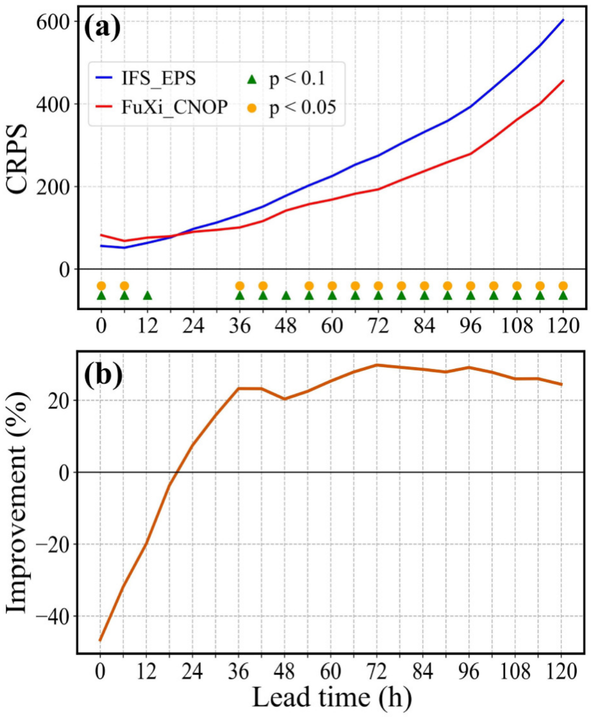

Fig. 3. Evaluation of the CRPS results. (a) CRPS comparison for 91 cases between FuXi_CNOP (red) and IFS_EPS (blue). Yellow circles (95% confidence) and green triangles (90% confidence) indicate statistically significant differences (Student's *t*-test). (b) Percentage improvement of FuXi_CNOP CRPS relative to IFS_EPS.

(see Fig. S7 in the ESM). This indicates that FuXi_CNOP achieves comparable reliability to IFS_EPS in probabilistic strike forecasts, as reflected by similar BS values, but its slightly inferior ROC performance highlights a trivially reduced discrimination capability. In other words, while FuXi_CNOP tends to assign probabilities that match IFS_EPS' strike frequencies (see Fig. 4 for some examples), its ability to more clearly separate strike from non-strike cases is somewhat weaker than that of IFS_EPS, resulting in less distinction between high- and low-risk scenarios. This limitation likely stems from the smoothing effect inherent in the MSE (mean squared error) training objective, which encourages a regression toward the mean and lacks the forecast sharpness required to distinguish between distinctly separate strike and non-strike scenarios.

In summary, FuXi_CNOP demonstrates significantly superior performance compared to IFS_EPS in deterministic forecasts across four key metrics: ensemble mean accuracy, ensemble spread consistency, ATE, and CTE. While the two systems exhibit comparable reliability in TC track strike probability forecasts—with FuXi_CNOP showing marginally lower discrimination ability—the former achieves more skillful probabilistic characterization of TC track evolution at extended lead times. Consequently, FuXi_CNOP generally outperforms IFS_EPS in both deterministic and probabilistic TC track forecasting, especially for long-range lead times.

3.3. *Mechanism*

Building on the deterministic and probabilistic forecast results of FuXi_CNOP with respect to TC tracks, this section elucidates the underlying mechanism that enables superior forecasting skill of TC tracks. By observing the first three CNOPs of each TC, it is found that their horizontal structures reveal their exceptional capability in identifying key sensitive areas responsible for TC track forecasting uncertainty. Figure 5 shows representative examples of the CNOPs for TC CHANTHU in the Northwest Pacific and THETA in the North Atlantic. For TC CHANTHU, its perturbation energies characterized by the CNOPs are found to predominantly concentrate in the spiral rainbands of the TC (see Fig. 5a), which often directly interface with the subtropical high periphery, consequently determining the TC track. In another case, for TC THETA in the North Atlantic (see Fig. 5b), its perturbation energies display a dual-core structure, with particularly pronounced perturbation maxima: the primary maximum is located within the southern subtropical high interaction zone, which shows strong dynamical coupling with the TC's subsequent northward track deflection; and the secondary maximum is situated in the eyewall region, with the asymmetry of the eyewall modulating the evolution of the TC wind field and, in turn, exerting an influence on the TC track (Cha et al., 2021). It is obvious that the CNOP patterns identify the dynamical instability within the TC region that capture the environmental modulation of TC circulation fields, consequently enhancing the subsequent TC track forecast skill—a finding consistent with previous studies (Zhou and Mu, 2011; Zhang et al., 2023).

Next, we adopt the statistics of 91 cases to elucidate the vertical energy distributions of the CNOPs (see Fig. 5c). The results show that the kinetic energy generally peaks at 400–300 hPa, coinciding with the TC warm-core structure (Li et al., 2025) and reflecting crucial baroclinic energy conversion processes (see Fig. S8 in the ESM), while internal energy maximizes in the 925–700 hPa boundary layer, corresponding to moisture transport (Li and Wang, 2021) and latent heat release mechanisms. The superimposed profile in Fig. 5c reveals a dual-peak vertical structure in energy distribution, demonstrating that the FuXi_CNOP model successfully captures both mid-upper level baroclinic processes and lower-level moist convective processes—the two primary physical mechanisms governing TC track uncertainty. Furthermore, the vertical energy distribution patterns of CNOPs for TCs in the western North Pacific and North Atlantic basins exhibit remarkable similarity (see Fig. S9 in the ESM), highlighting consistent dynamical behaviors across these geographically distinct regions.

As indicated by the CNOP structures, the CNOPs effectively identify the key areas responsible for TC track forecasting uncertainties and capture both mid-upper level baroclinic processes and lower-level moist convective processes governing TC track uncertainty. We therefore have strong reason to believe that the ensemble members generated by FuXi_CNOP are capable of representing the TC track fore-

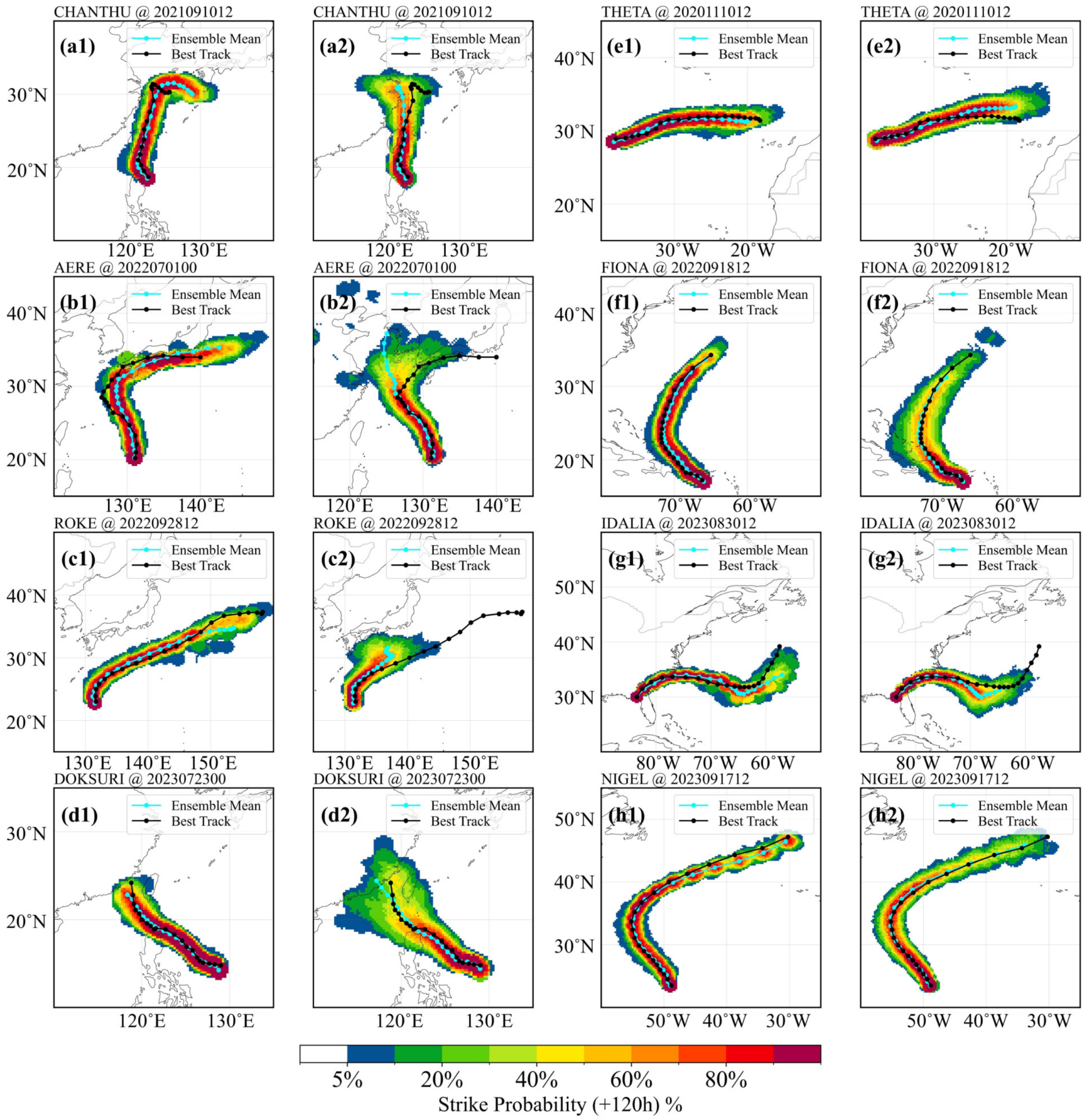

Fig. 4. TC strike probability forecasts for TCs (a–d) CHANTHU, AERE, ROKE, and DOKSURI in the Northwest Pacific, and (e–h) THETA, FIONA, IDALIA, and NIGEL in the North Atlantic. The suffix "1" denotes FuXi_CNOP, while "2" represents IFS_EPS. Forecasts include the best track (black line), ensemble mean (blue line), and strike probability density (shaded).

casting uncertainty and provide reliable forecasts of TC tracks. It is known that TC motion is primarily modulated by the steering flow related to subtropical high-pressure systems (Wu et al., 2005; Kossin et al., 2010; Zhang et al., 2023). We plot in Fig. 6 the ensemble spread of the wind generated by FuXi_CNOP and IFS_EPS for the TCs as in Fig. 4. It is shown that FuXi_CNOP demonstrates superior dynamical consistency in representing the steering flow interactions, as evidenced by the coherent spatial coupling between its 850–250 hPa mean wind fields and wind speed spread pattern. Specifically, for the TC cases of CHANTHU and THETA (see Figs. 6a1 and 6e1; consistent with the case selections in Fig. 5), it is illustrated that the wind speed spread exhibits a clear correspondence with the sensitive areas and is well coupled with the TC steering flow, demonstrating that the CNOPs effectively capture the uncertainties that influence TC track evolution. Further, we adopt TC AERE to exemplify this advantage (see Fig. 6b1): the system not only accurately resolves the TC core structure, known to critically influence TC development (Chen et al., 2018; Qin et al., 2020), but also optimally localizes maximum spread regions along the interaction areas on the southwestern side of the TC, where the southwesterly steering flow (associated with the subtropical high) interacts with the TC circulation—pre-

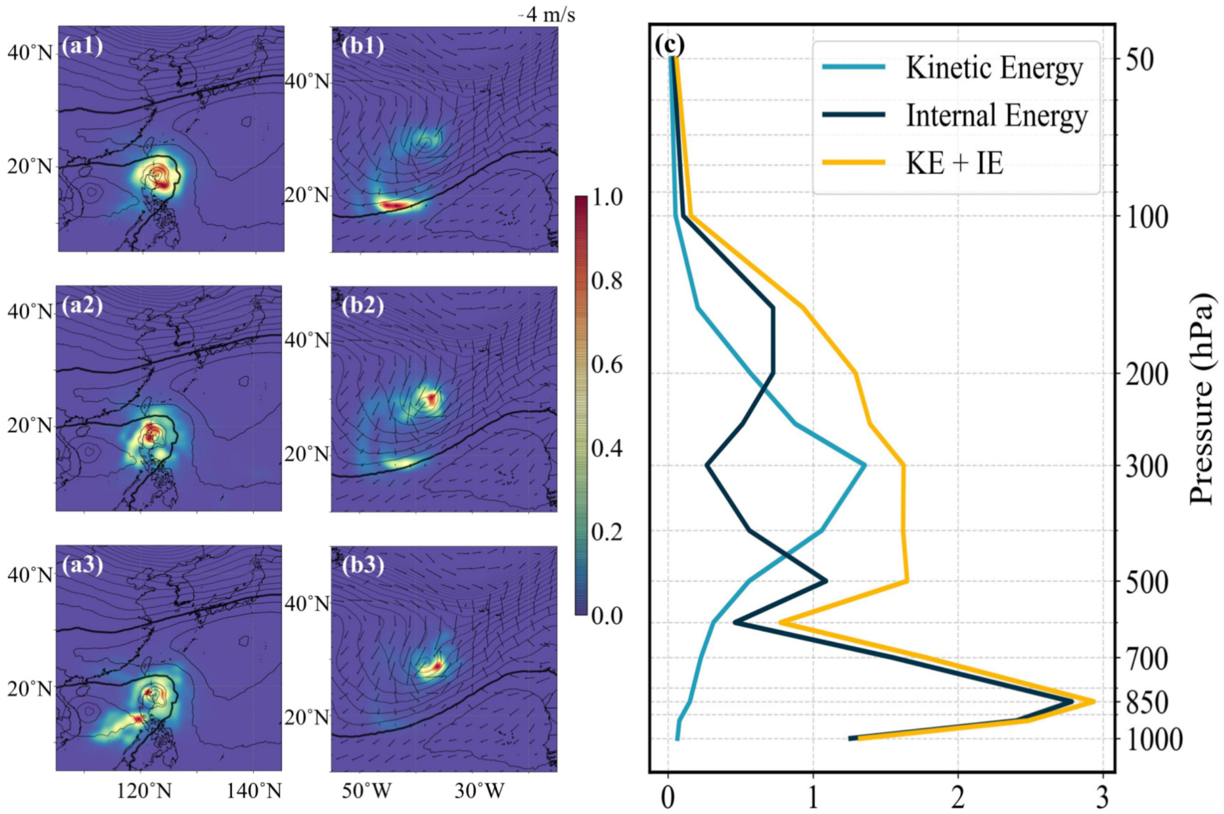

Fig. 5. Horizontal and vertical perturbation structures of CNOPs. (a) TC CHANTHU (1200 UTC 10 September 2021) in the Northwest Pacific and (b) TC THETA (1200 UTC 10 November 2020) in the North Atlantic, showing vertically integrated CNOP energy (shaded, normalized 0–1) with geopotential height contours (black lines; thick line indicates 5880 gpm) and 10-m wind (black arrows). Labels 1, 2, and 3 denote the 1st, 2nd, and 3rd CNOP pattern, respectively. (c) Mean horizontally integrated energy profiles for 91 cases: kinetic energy (sky blue), internal energy (dark blue), and their sum (yellow).

cisely corresponding to the primary sources of track forecast uncertainty. In contrast, IFS_EPS exhibits spread maxima not only in the TC vicinity but also prominently in the South China Sea monsoon region (15°–20°N, 110°–120°E) (see Fig. 6b2). This dual-maxima pattern indicates that the system's uncertainty representation is not optimally focused on TC-specific dynamics, instead maintaining broadly distributed spread across multiple synoptic systems. These findings demonstrate that FuXi_CNOP enables more precise identification and representation of key uncertainty sources in TC track forecasting, thus providing an interpretation of the superior performance of FuXi_CNOP in TC track forecasting.

4. Discussion

This study introduces a novel AI-driven ensemble forecasting framework (FuXi_CNOP) that synergizes dynamical O-CNOPs optimization with advanced meteorological modeling. By integrating the FuXi AI large model, the system achieves unprecedented computational efficiency while maintaining rigorous dynamical consistency. When specifically applied to TC track forecasting, the FuXi_CNOP framework demonstrates comprehensive superiority over the benchmark IFS_EPS system: it exhibits enhanced ensemble mean accuracy, superior ensemble spread consistency, and significant reductions in both along-track and cross-track errors. While both systems provide comparable reliability in probabilistic strike forecasts, FuXi_CNOP establishes a distinct advantage in deterministic track forecasts and generally excels in probabilistic characterization of TC evolution at extended lead times. Physically, the framework's perturbation energy structures reveal distinct spatial patterns: horizontal energy concentrates in TC core regions and subtropical high interface zones, while vertical energy shows dual-peak distributions corresponding to baroclinic and moist convective processes; these structural characteristics enable FuXi_CNOP to more effectively capture critical environmental factors governing TC motion, particularly through improved representation of subtropical high–TC circulation interactions. Consequently, the framework delivers more reliable probability forecasts for complex track variations, establishing itself as a transformative solution for operational weather prediction, especially in long-range forecasting scenarios.

From an alternative perspective, end-to-end AI-driven ensemble forecasting systems have been developed (Chen et al., 2024; Zhong et al., 2024b; Price et al., 2025; Lang et al., 2026). These systems directly generate ensemble mem-

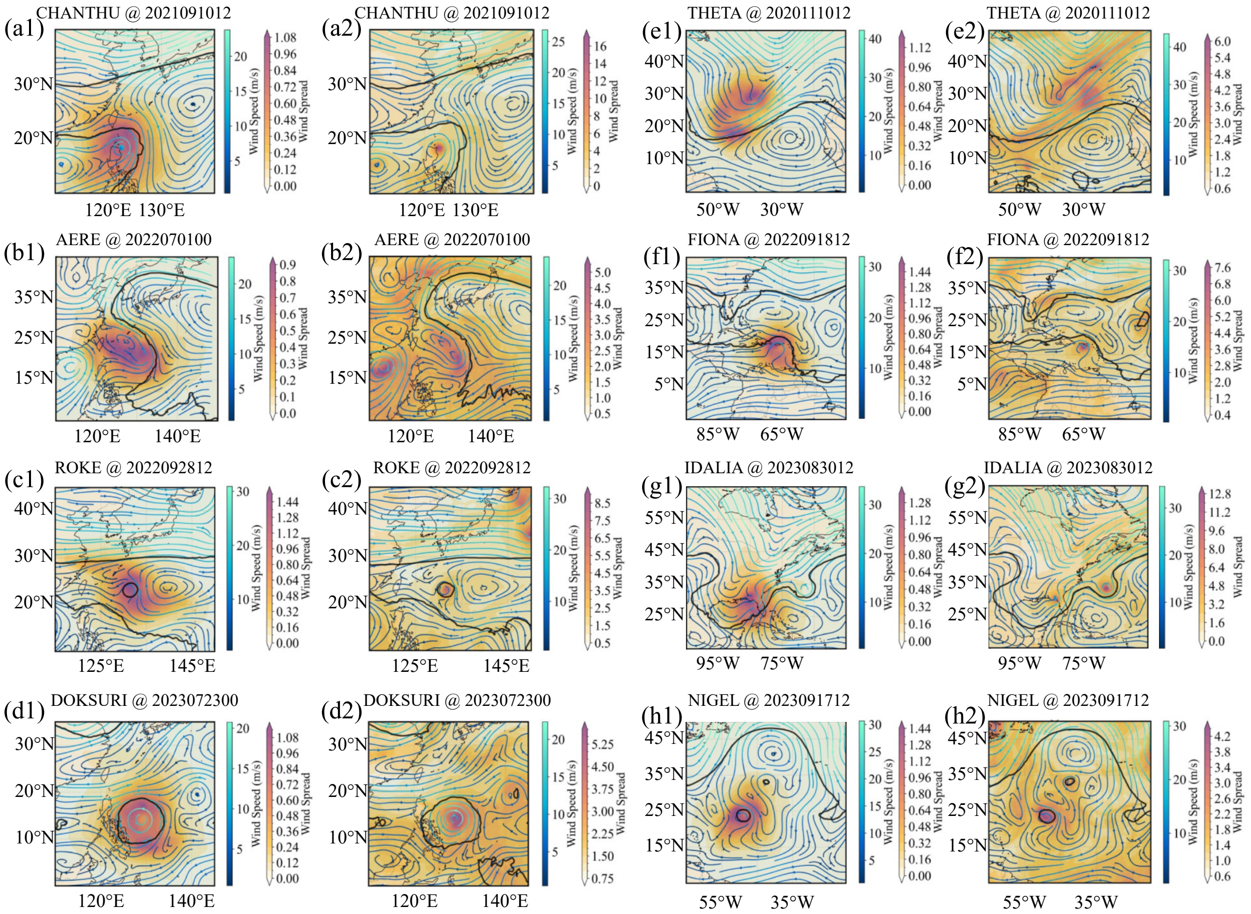

Fig. 6. Ensemble-mean deep-layer (850–250 hPa) wind fields (streamlines) and associated wind speed spread (shading) at the initial time. Figures (a–d) are for TCs CHANTHU, AERE, ROKE, and DOKSURI in the Northwest Pacific, and (e–h) are for TCs THETA, FIONA, IDALIA, and NIGEL in the North Atlantic, with the suffix "1" denoting FuXi_CNOP results and "2" representing IFS_EPS outputs. The 5880 gpm geopotential height contour at 500 hPa is overlaid in black.

bers using generative AI models. While such AI-driven ensembles, when using ERA5 reanalysis data as inputs, bypass the need to consider initial error effects, this approach remains impractical for operational forecasting for two key reasons: (1) real-time reanalysis data are unavailable during actual forecasting operations; and (2) forecast uncertainties would theoretically require an infinite number of members for complete characterization. Additionally, the interpretability of these ensemble members is limited by the latent-space representation of variables. Pu et al. (2025) attempted to address this by generating ensembles through initial perturbations applied to a deterministic control forecast produced by an AI model. However, their methodology inherits similar limitations: their control forecasts depend on the unrealistic assumption of ERA5 reanalysis data availability, and the perturbations are generated using a lag-based method. Furthermore, their approach demands an impractically large ensemble size (e.g., 2000 members generated in their study) to achieve reliable forecasts. In contrast, the FuXi_CNOP system represents a significant advancement by adopting a forecasting field as the initial input for control forecast generation. This approach is highly applicable to operational forecasting. Remarkably, this system achieved superior ensemble forecast skill in TC track forecasts compared to the IFS_EPS system, despite using a substantially smaller ensemble size. This innovation marks a crucial step toward operational AI-based forecasting, though further research is needed to fully realize its operational potential.

While FuXi_CNOP is reliable in early stages, its ensemble spread lags behind RMSEs at extended lead times. This under-dispersion may stem from the smoothing effect induced by training the model to minimize the MSE, which limits the growth of ensemble spread and underestimates forecast uncertainties during long-term iterative rollouts. Addressing this challenge, a nonlinear forcing singular vector method (C-NFSV) (Duan et al., 2022) offers a promising framework. By dynamically coordinating initial and model perturbations, this approach could potentially compensate for the suppressed growth and substantially enhance the reliability of AI-driven ensemble forecasts. When applied to AI models, this approach may not only mitigate error propagation but also unlock synergistic improvements by dynamically weighting initial and model uncertainties. In any case, for AI models exhibiting chaotic behavior, the above ensemble strategy will be particularly promising. By generating fast-growing perturbations on control forecasts generated by AI models, it could offer a transformative pathway for next-generation ensemble forecasting systems.

Acknowledgments. This study was supported by funding from the National Natural Science Foundation of China (Grant Nos. 42330111; 42505065). We are grateful for the technical support of the National Large Scientific and Technological Infrastructure "Earth System Numerical Simulation Facility" (https://cstr.cn/31134.02.EL).

Data availability statement. ERA5 data are available at https://cds.climate.copernicus.eu/cdsapp#!/search?type=dataset; TIGGE data are available at https://apps.ecmwf.int/datasets/data/tigge/levtype=sfc/type=cf/; the IBTrACS dataset is available at https://www.ncei.noaa.gov/products/international-best-track-archive. The FuXi model is open source and available at https://github.com/tpys/FuXi.

Electronic supplementary material. Supplementary material is available in the online version of this article at https://doi.org/10.1007/s00376-026-5784-3.

APPENDIX A

To enable a quantitative evaluation, metrics for both deterministic and probabilistic forecast skills are presented. The TC track forecast error is defined as the great-circle distance between the forecasted and best-track TC centers (Cox et al., 2018), which is expressed as follows:

$$d = 2R\arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\varphi}{2}\right)+\cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right), \tag{A1}$$

where $\Delta\varphi = \varphi_2-\varphi_1$, $\Delta\lambda = \lambda_2-\lambda_1$, R = 6371 km (Earth's radius), and (φ_1,λ_1) and (φ_2,λ_2) denote the latitudes and longitudes of the two points. Each TC track error is further decomposed into ATE and CTE components (WMO, 2013; Heming, 2017): positive CTE indicates that the forecasted track lies to the right of the best track, and positive ATE indicates that the forecasted TC moves faster than indicated by the best track.

The ensemble mean track represents the average position across all ensemble members, which is calculated as follows:

$$\left(\overline{\lambda},\overline{\varphi}\right) = \frac{1}{N}\sum_{i=1}^{N}(\lambda_i,\varphi_i)\ , \tag{A2}$$

where λ_i,φ_i are the longitude and latitude of the ith ensemble member, and N is the number of ensemble members. The ensemble spread is defined as the standard deviation of members relative to the mean, which can be calculated as follows:

$$\text{Spread} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\mathbf{F}_i-\overline{\mathbf{F}}\right|^2}\ , \tag{A3}$$

where $\mathbf{F}_i$ is the track of the ith ensemble member, $\overline{\mathbf{F}}$ is the ensemble mean, and $\left|\mathbf{F}_i-\overline{\mathbf{F}}\right|$ is the great-circle distance (for tracks) or wind speed difference (for wind fields). A perfect ensemble forecast is proven to have a relationship in which the ensemble spread is equal to the ensemble mean forecasting error (Bowler, 2006; Hopson, 2014).

The CRPS measures the distance between forecasted and observed TC distributions (Gneiting and Raftery, 2007; Székely and Rizzo, 2013):

$$\text{CRPS} = 2\text{E}|\mathbf{X}-\mathbf{Y}|-\text{E}|\mathbf{X}-\mathbf{X}'|-\text{E}|\mathbf{Y}-\mathbf{Y}'|\ . \tag{A4}$$

In our research, $\mathbf{X}$ is the 2-dimensional TC location from the FuXi model forecast; $\mathbf{Y}$ is the TC location from IBTrACS data; $\mathbf{X}'$ and $\mathbf{Y}'$ are independent and identically distributed copies of $\mathbf{X}$ and $\mathbf{Y}$. E represent mathematical expectation and $|\cdot|$ represent the great-circle distance metric (physically more appropriate).

The BS is the mean squared error of the probability forecasts, defined as follows:

$$\text{BS} = \frac{1}{N}\sum_{i=1}^{N}(f_i-o_i)^2 \tag{A5}$$

where N is the number of realizations of the prediction processed, and f_i and o_i are the probabilities of the forecast and observation for the prediction processes, respectively. A smaller BS indicates a better probability forecast skill (Brier, 1950).

The ROC measures the ability of a forecast to discriminate between events and nonevents. This curve is a function of the hit rate and false alarm rate. Thus, higher skill is indicated by a ROC that is closer to the top-left corner of the diagram (implying a low false alarm rate and high hit rate) or a larger ROCA (Mason and Graham, 2002).

REFERENCES

Bi, K. F., L. X. Xie, H. H. Zhang, X. Chen, X. T. Gu, and Q. Tian, 2023: Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, **619**, 533–538, https://doi.org/10.1038/s41586-023-06185-3.

Bourke, W., T. Hart, P. Steinle, R. Seaman, G. Embery, M. Naughton, and L. Rikus, 1995: Evolution of the Bureau of Meteorology's global assimilation and prediction system. Part 2: Resolution enhancements and case studies. *Aust. Meteor. Mag.*, **44**, 19–40, https://doi.org/10.1071/ES95002.

Bowler, N. E., 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus A: Dynamic Meteorology and Oceanography*, **58**, 538–548, https://doi.org/10.1111/j.1600-0870.2006.00197.x.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, https://doi.org/10.

1175/1520-0493(1950)078<0001:vofeit>2.0.co;2.

Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456, https://doi.org/10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2.

Cha, T.-Y., Bell, M. M., and DesRosiers, A. J., 2021: Doppler Radar Analysis of the Eyewall Replacement Cycle of Hurricane Matthew (2016) in Vertical Wind Shear. *Mon. Wea. Rev.*, **149**, 2927–2943, https://doi.org/10.1175/MWR-D-20-0289.1.

Chen, G. H., C.-C. Wu, and Y.-H. Huang, 2018: The role of near-core convective and stratiform heating/cooling in tropical cyclone structure and intensity. *J. Atmos. Sci.*, **75**, 297–326, https://doi.org/10.1175/JAS-D-17-0122.1.

Chen, L., X. H. Zhong, F. Zhang, Y. Cheng, Y. H. Xu, Y. Qi, and H. Li, 2023: FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. *npj Climate and Atmospheric Science*, **6**, 190, https://doi.org/10.1038/s41612-023-00512-1.

Chen, L., and Coauthors, 2024: A machine learning model that outperforms conventional global subseasonal forecast models. *Nature Communications*, **15**, 6425, https://doi.org/10.1038/s41467-024-50714-1.

Conroy, A., and Coauthors, 2023: Track forecast: Operational capability and new techniques - Summary from the Tenth International Workshop on Tropical Cyclones (IWTC-10). *Tropical Cyclone Research and Review*, **12**, 64–80, https://doi.org/10.1016/j.tcrr.2023.05.002.

Cox, T. S., C. S. H. Hoi, C. K. Leung, and C. R. Marofke, 2018: An accurate model for hurricane trajectory prediction. *Proc. 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, Tokyo, Japan, IEEE, 534–539, https://doi.org/10.1109/COMPSAC.2018.10290.

Duan, W. S., and Z. H. Huo, 2016: An approach to generating mutually independent initial perturbations for ensemble forecasts: Orthogonal conditional nonlinear optimal perturbations. *J. Atmos. Sci.*, **73**, 997–1014, https://doi.org/10.1175/JAS-D-15-0138.1.

Duan, W. S., J. J. Ma, and S. Vannitsem, 2022: An ensemble forecasting method for dealing with the combined effects of the initial and model errors and a potential deep learning implementation. *Mon. Wea. Rev.*, **150**, 2959–2976, https://doi.org/10.1175/mwr-d-22-0007.1.

Duan, W. S., L. C. Yang, Z. Z. Xu, and J. Chen, 2023: Conditional nonlinear optimal perturbation: Applications to ensemble forecasting of high-impact weather systems. *Numerical Weather Prediction: East Asian Perspectives*, S. K. Park, Ed., Springer, 441–460, https://doi.org/10.1007/978-3-031-40567-9_17.

Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378, https://doi.org/10.1198/016214506000001437.

Gori, A., N. Lin, D. Z. Xi, and K. Emanuel, 2022: Tropical cyclone climatology change greatly exacerbates US extreme rainfall–surge hazard. *Nature Climate Change*, **12**, 171–178, https://doi.org/10.1038/s41558-021-01272-7.

Heming, J. T., 2017: Tropical cyclone tracking and verification techniques for Met Office numerical weather prediction models. *Meteorological Applications*, **24**, 1–8, https://doi.org/10.1002/met.1599.

Hopson, T. M., 2014: Assessing the ensemble spread–error relationship. *Mon. Wea. Rev.*, **142**, 1125–1142, https://doi.org/10.1175/MWR-D-12-00111.1.

Houtekamer, P. L., L. Lefaivre, and J. Derome, 1996: The RPN ensemble prediction system. *Proc. ECMWF Seminar on Predictability*, Reading, UK, ECMWF, 121–146.

Huo, Z. H., and W. S. Duan, 2019: The application of the orthogonal conditional nonlinear optimal perturbations method to typhoon track ensemble forecasts. *Science China Earth Sciences*, **62**, 376–388, https://doi.org/10.1007/s11430-018-9248-9.

Kossin, J. P., S. J. Camargo, and M. Sitkowski, 2010: Climate modulation of North Atlantic hurricane tracks. *J. Climate*, **23**, 3057–3076, https://doi.org/10.1175/2010JCLI3497.1.

Lam, R., and Coauthors, 2023: Learning skillful medium-range global weather forecasting. *Science*, **382**, 1416–1421, https://doi.org/10.1126/science.adi2336.

Lang, S., and Coauthors, 2024: AIFS – ECMWF's data-driven forecasting system. arXiv: 2406.01465, https://doi.org/10.48550/arXiv.2406.01465.

Lang, S., and Coauthors, 2026: AIFS-CRPS: Ensemble forecasting using a model trained with a loss function based on the continuous ranked probability score. *npj Artificial Intelligence*, **2**, 18, https://doi.org/10.1038/s44387-026-00073-7.

Li, T.-H., and Y. Q. Wang, 2021: The role of boundary layer dynamics in tropical cyclone intensification. Part II: Sensitivity to initial vortex structure. *J. Meteor. Soc. Japan*, **99**, 555–573, https://doi.org/10.2151/jmsj.2021-028.

Li, Y. H., W. S. Duan, W. Han, H. Li, and X. H. Qin, 2025: Improving tropical cyclone track forecast skill through assimilating target observation achieved by AI-based conditional nonlinear optimal perturbation. *J. Geophys. Res.*, **130**, e2024JD043261, https://doi.org/10.1029/2024JD043261.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, https://doi.org/10.1111/j.2153-3490.1969.tb00444.x.

Mason, S. J., and N. E. Graham, 2002: Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quart. J. Roy. Meteor. Soc.*, **128**, 2145–2166, https://doi.org/10.1256/003590002320603584.

Mureau, R., F. Molteni, and T. N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.*, **119**, 299–323, https://doi.org/10.1002/qj.49711951005.

Nystrom, R. G., F. Q. Zhang, E. B. Munsell, S. A. Braun, J. A. Sippel, Y. H. Weng, and K. Emanuel, 2018: Predictability and dynamics of Hurricane Joaquin (2015) explored through convection-permitting ensemble sensitivity experiments. *J. Atmos. Sci.*, **75**, 401–424, https://doi.org/10.1175/JAS-D-17-0137.1.

Paszke, A., and Coauthors, 2017: Automatic differentiation in PyTorch. *Proc. 31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, Curran Associates Inc.

Peduzzi, P., B. Chatenoux, H. Dao, A. De Bono, C. Herold, J. Kossin, F. Mouton, and O. Nordbeck, 2012: Global trends in tropical cyclone risk. *Nature Climate Change*, **2**, 289–294, https://doi.org/10.1038/nclimate1410.

Price, I., and Coauthors, 2025: Probabilistic weather forecasting with machine learning. *Nature*, **637**, 84–90, https://doi.org/10.1038/s41586-024-08252-9.

Pu, J. C., M. Mu, J. Feng, X. H. Zhong, and H. Li, 2025: A fast physics-based perturbation generator of machine learning weather model for efficient ensemble forecasts of tropical cyclone track. *npj Climate and Atmospheric Science*, **8**, 128, https://doi.org/10.1038/s41612-025-01009-9.

Qin, X. H., W. S. Duan, and H. Xu, 2020: Sensitivity to tendency perturbations of tropical cyclone short-range intensity forecasts generated by WRF. *Adv. Atmos. Sci.*, **37**, 291–306, https://doi.org/10.1007/s00376-019-9187-6.

Székely, G. J., and M. L. Rizzo, 2013: Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, **143**, 1249–1272, https://doi.org/10.1016/j.jspi.2013.03.018.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, https://doi.org/10.1175/1520-0477(1993)074<2317:efantg>2.0.co;2.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319, https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.

Wang, Y., and C.-C. Wu, 2004: Current understanding of tropical cyclone structure and intensity changes – a review. *Meteor. Atmos. Phys.*, **87**, 257–278, https://doi.org/10.1007/s00703-003-0055-6.

WMO, 2013: Verification methods for tropical cyclone forecasts. Available from https://www.semanticscholar.org/paper/Verification-methods-for-tropical-cyclone-forecasts-Forecasts/127a6d9a6322759b1ae692b92d95d67c86f2eb57.

Wu, L. G., B. Wang, and S. Q. Geng, 2005: Growing typhoon influence on east Asia. *Geophys. Res. Lett.*, **32**, L18703, https://doi.org/10.1029/2005GL022937.

Zhang, H., W. S. Duan, and Y. C. Zhang, 2023: Using the orthogonal conditional nonlinear optimal perturbations approach to address the uncertainties of tropical cyclone track forecasts generated by the WRF model. *Wea. Forecasting*, **38**, 1907–1933, https://doi.org/10.1175/WAF-D-22-0175.1.

Zhang, Y., G. F. Fan, Y. He, and L. J. Cao, 2017: Risk assessment of typhoon disaster for the Yangtze River Delta of China. *Geomatics, Natural Hazards and Risk*, **8**, 1580–1591, https://doi.org/10.1080/19475705.2017.1362040.

Zhong, X. H., L. Chen, J. Liu, C. S. Lin, Y. Qi, and H. Li, 2024a: FuXi-Extreme: Improving extreme rainfall and wind forecasts with diffusion model. *Science China Earth Sciences*, **67**, 3696–3708, https://doi.org/10.1007/s11430-023-1427-x.

Zhong, X. H., and Coauthors, 2024b: FuXi-ENS: A machine learning model for medium-range ensemble weather forecasting. arXiv: 2405.05925, https://doi.org/10.48550/arXiv.2405.05925.

Zhou, F. F., and M. Mu, 2011: The impact of verification area design on tropical cyclone targeted observations based on the CNOP method. *Adv. Atmos. Sci.*, **28**, 997–1010, https://doi.org/10.1007/s00376-011-0120-x.