Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

FuXi-TC: A generative framework integrating deep learning and physics-based models for improved tropical cyclone forecasts

Shan Guo
Shanghai Academy of Artificial Intelligence for Science
Lei Chen
Fudan University
Yangyang Zhao
Fudan University
Yuetan Lin
Shanghai Academy of Artificial Intelligence for Science
Zeyi Niu
Fudan University
Xinyan Zhang
China Meteorological Administration
Ziyao Sun
China Meteorological Administration
Xiaohui Zhong
Fudan University
Hao Li
lihao_lh@fudan.edu.cn

Fudan University

Article

Keywords: Tropical Cyclone Intensity, FuXi, Diffusion model, WRF

Posted Date: April 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9415250/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at npj Climate and Atmospheric Science on September 8th, 2026. See the published version at https://doi.org/10.1038/s41612-026-01501-w.

FuXi-TC: A generative framework integrating deep learning and physics-based models for improved tropical cyclone forecasts

Shan Guo[2†], Lei Chen[1,2†], Yangyang Zhao[1,2], Yuetan Lin[1], Zeyi Niu[3,4], Xinyan Zhang[3], Ziyao Sun[3,4], Xiaohui Zhong[1,2*], Hao Li[1,2*]

[1]Artificial Intelligence Innovation and Incubation Institute, Fudan University, Shanghai, 200433, China.
[2]Shanghai Academy of Artificial Intelligence for Science, Shanghai, 200232, China.
[3]Department of Atmospheric and Oceanic Sciences, Institute of Atmospheric Sciences, Fudan University, Shanghai, 200433, China.
[4]Shanghai Typhoon Institute, China Meteorological Administration, Shanghai, 200030, China.

*Corresponding author(s). E-mail(s): x7zhong@gmail.com; lihao_lh@fudan.edu.cn;
Contributing authors: guoshan@sais.org.cn; cltpys@163.com; zhaoyangyang@sais.org.cn; linyuetan@sais.org.cn; niuzy@typhoon.org.cn; zhangxy@typhoon.org.cn; sunzy@typhoon.org.cn;
†These authors contributed equally to this work.

Abstract

Tropical cyclones (TCs) are among the most devastating natural hazards, yet their intensity remains notoriously difficult to predict. Numerical weather prediction (NWP) models are constrained by both computational demands and intrinsic predictability, while state-of-the-art deep learning-based weather forecasting models tend to underestimate TC intensity due to biases in reanalysis-based training data.Here, we present FuXi-TC, a diffusion-based generative forecasting framework that combines the track prediction strength of the FuXi

model with the intensity representation of NWP simulations. By conditioning a diffusion model on the large-scale forecasts of the global FuXi model, FuXi-TC effectively downscales and delivers higher-accuracy forecasts of fine-grained variable fields such as wind speed and precipitation. In evaluations across the 2024 Western North Pacific, our approach matches the TC intensity forecast skill of the operational European Centre for Medium-Range Weather Forecasts (ECMWF) deterministic model while delivering superior precipitation forecasts. Meanwhile this is achieved with significantly higher inference speeds and lower computational costs. Moreover, FuXi-TC demonstrates robust zero-shot generalization directly when applied to North Atlantic hurricanes without any fine-tuning or re-training. When applied to the FuXi ensemble model, this framework effectively yields well-dispersed probabilistic forecasts and refines the ensemble intensity predictions. More broadly, these results highlight the potential of leveraging physics-based simulations to construct new training datasets that further advancing deep learning–based weather forecasting.

Keywords: Tropical Cyclone Intensity, FuXi, Diffusion model, WRF

1 Introduction

Tropical cyclones (TCs) frequently inflict catastrophic damage on coastal and inland regions through severe winds, heavy rainfall, storm surges, and widespread flooding. They pose serious threats to infrastructure, economic activities, and human safety across affected areas [1–5]. Under global warming, although the global annual frequency of TCs has remained relatively stable, the occurrence and spatial extent of extreme-intensity events have increased [6]. Continued ocean warming is expected to further intensify TC strength, enhance precipitation, and prolong cyclone lifespans, amplifying societal risk and increasing demands on disaster prevention and mitigation systems [7].

Accurate and timely prediction of TC track and intensity remains a critical challenge for modern meteorological services. Operational forecasts primarily rely on traditional numerical weather prediction (NWP) models specifically optimized for tropical cyclones. These include the Hurricane Weather Research and Forecasting (HWRF) model [8, 9] from the National Oceanic and Atmospheric Administration (NOAA), Hurricane Analysis and Forecast System (HAFS) [10] from the National Centers for Environmental Prediction (NCEP), and China Meteorological Administration Typhoon Model (CMA-TYM) [11]. These models solve atmospheric dynamical and thermodynamical equations to simulate the evolution of large-scale circulation and weather systems from prescribed initial and boundary conditions. However, their forecast accuracy is constrained by uncertainties in initial conditions, errors in physical parameterizations, and limited spatial resolution. Furthermore, high-resolution NWP simulations are computationally expensive, typically requiring several hours to complete, which limits their utility for rapid-response forecasting during extreme

weather events. Recent studies suggest that the predictability of TC track and intensity based on traditional NWP models may be approaching theoretical limits [12, 13], with marginal returns from further increase in resolution.

In recent years, data-driven weather forecasting models have demonstrated strong skill in TC track prediction [14]. For instance, Pangu-Weather [15] outperformed the high-resolution forecasts (HRES) of the European Centre for Medium-Range Weather Forecasts (ECMWF) in track prediction accuracy for 88 TCs in 2018 . Similarly, FuXi-Extreme [16], evaluated across 90 TCs, achieved superior performance in both TC track prediction and large-scale circulation patterns in influencing TC motion, particularly at lead times beyond 48 hours. These systems offer fast inference and scalability, making them attractive for real-time applications [17, 18]. Despite these advances, TC intensity prediction remains a key limitation of current data-driven models [13]. Artificial intelligence (AI) models are typically optimized for global statistical performance and tend to smooth extremes, leading to systematic underestimation of maximum winds [16, 19, 20]. This issue is compounded by dependence on reanalysis datasets such as ERA5 for training [21], which themselves underrepresent TC intensity due to resolution constraints and assimilation processes [22, 23]. Efforts to improve intensity forecasts using auxiliary datasets, such as International Best Track Archive for Climate Stewardship (IBTrACS) [24], have primarily focused on predicting scalar attributes such as maximum sustained wind speed or minimum central pressure. While valuable, these approaches do not directly enhance the full spatiotemporal structure of the TCs, which is critical for hazards such as wind damage and rainfall distribution. Wang et al. [25] improved long-term tropical cyclone intensity forecasting by jointly leveraging IBTrACS intensity records and ERA5 reanalysis data. The model learns a discrete representation of historical intensity conditioned on surrounding ERA5 atmospheric fields, and performs temporal prediction in this latent space. Forecasts from the FengWu model are incorporated as conditioning information, and potential intensity calculated from the forecast fields is introduced to mitigate systematic underestimation and reduce long-term error accumulation. Similarly, Huang et al. [26] constructed a global multimodal dataset by aligning best-track intensity records with meteorological reanalysis fields and environmental variables. By explicitly modeling environmental contexts such as the subtropical high, this method significantly improves track and intensity predictions. Although both frameworks ingest complex multi-dimensional spatial meteorological grids as inputs, their final predictions are limited to time-series sequences of TC intensity. Consequently, these models fail to forecast the full spatiotemporal structure of the TC, such as the asymmetric wind fields or precipitation distributions, which is critical for accurately assessing localized hazards like wind damage and storm surges.

Hybrid strategies have attempted to integrate AI model forecasts with physical models to improve TC intensity prediction. These approaches leverage the superior large-scale predictive skill of AI models to provide improved initial and boundary conditions, while relying on physics-based NWP models to simulate fine-scale intensity evolution and spatiotemporal structures. Xu et al.[13] proposed a novel framework that uses the Pangu-Weather forecasts to drive the high-resolution Weather Research and Forecasting (WRF) model, augmented with vortex dynamic initialization. This

approach improves typhoon intensity forecasts and demonstrates the promising potential of AI models to serve as superior large-scale forcings for physics-based regional forecasting. Liu et al. [20] developed a novel hybrid forecasting framework that downscales Pangu-Weather forecasts with a high-resolution WRF model, incorporating large-scale spectral nudging and ocean mixed-layer coupling to extend tropical cyclone prediction to two weeks. This framework advances extended-range TC intensity and structural forecasting capability compared with AI models. Syed et al. [27] proposes a hybrid NWP-AI forecasting system that employs large-scale spectral nudging to constrain the physics-based Global Environmental Multiscale model toward the superior large-scale predictions of the GraphCast model. This approach successfully leverages the strengths of both paradigms, and resulting in improved forecast skill across variables and lead times. However, such hybrid approaches still depend on regional high-resolution NWP model for dynamical downscaling. The high computational cost and slow integration speed of traditional NWP remain major bottlenecks for operational deployment. A method that improves TC intensity and structure while preserving the computational efficiency of pure data-driven models is still lacking.

In this study, we present a fully data-driven framework for the rapid refinement of TC structure and rainfall forecasts, while preserving the inherent advantages of large meteorological models in accurate TC track prediction. Specifically, we introduce FuXi-TC, a lightweight downscaling and refinement module based on a denoising diffusion probabilistic model (DDPM) [28, 29] that enhances 5-day forecasts produced by FuXi [30]. Unlike the aforementioned hybrid approaches that require computationally expensive NWP models during the inference phase, our framework decouples the physical simulation from the real-time forecasting process. To address the intensity underestimation inherent in ERA5 and the limitation of IBTrACS-based methods that optimize scalar sequences, a key component of our approach is a physics-guided data construction strategy. By generating multi-variable meteorological fields using WRF simulations as training targets, this strategy enables the trained model to reconstruct and refine the entire spatiotemporal structure of the TC. Crucially, these high-resolution target fields are generated using WRF simulations dynamically downscaled from FuXi forecasts rather than from ERA5. This design preserves structural and spatial consistency between the input FuXi forecasts and target fields, minimizing spatial misalignment arising from track deviations. The resulting paired dataset facilitates the diffusion-based FuXi-TC model to focus on structural refinement and downscaling, rather than compensating for positional discrepancies. The proposed framework is lightweight and highly adaptable, allowing for robust cross-basin generalization and extension to other existing large meteorological models such as FuXi-ENS[31]. By combining physics-guided target generation with generative refinement, our approach advances TC intensity prediction while retaining the speed and scalability of AI models, offering a practical solution for operational deployment.

2 Results

For evaluating the comprehensive performance of the proposed FuXi-TC framework, our assessment encompasses several key dimensions. First, we analyze the overall deterministic forecast performance in the Western North Pacific (WNP), focusing on TC track, intensity, and precipitation metrics, with supplementary evaluations on mean sea-level pressure (MSL; Supplementary Fig. 8) and additional precipitation statistics (Supplementary Figs. 5 and 6) provided in the supplementary materials. Second, we present in-depth TC case studies of extreme events to illustrate the model's capability in capturing fine-scale wind structures and rainfall distribution. Third, we evaluate the probabilistic forecasting performance by integrating FuXi-TC into the FuXi-ENS, resulting in an enhanced ensemble framework termed FuXi-TC-ENS. We analyze its ensemble spread characteristics and intensity estimation skill to assess uncertainty representation and probabilistic reliability. Fourth, we demonstrate the model's zero-shot geographical generalization by extending the evaluation to the North Atlantic (NA) basin, supported by specific hurricane case studies and domain configurations detailed in the supplementary materials. Finally, we assess the computational efficiency of our approach relative to traditional NWP models. Furthermore, additional evaluations, such as physical consistency using kinetic energy spectra to verify the effective resolution and spatial variance of the model outputs, are provided in supplementary information (Supplementary Fig. 7).

2.1 Overall TC forecast performance

Evaluation of TC forecasts includes assessments of both track and intensity prediction. In this study, we have analyzed the forecast performance for 17 TCs (see Supplementary Table 1) that occurred during July–October 2024, which represents the typhoon season in the WNP. Fig. 1 presents a statistical comparison between ECMWF high-resolution deterministic forecast (HRES-0.1°), FuXi, WRF-0.1° and FuXi-TC-0.1° in 5-day forecasts as well as ERA5 reanalysis data against IBTrACS dataset. In terms of TC intensity forecasts, FuXi-TC-0.1° trained using WRF-0.1° presents a more noticeable improvement over FuXi and ERA5 in terms of the maximum 10-m wind speed (WS10M), as evidenced by lower MAE values for Vmax. Furthermore, FuXi-TC-0.1° demonstrates overall comparable performance to HRES-0.1°. Specifically, it outperforms HRES-0.1° during the first 48 hours of the forecast, although it performs marginally worse in the final 12 hours. Regarding TC track forecasts, FuXi, FuXi-TC-0.1°, and WRF-0.1° exhibit negligible differences in track errors, and all perform better compared to HRES-0.1° at lead times of 84-120 hours.

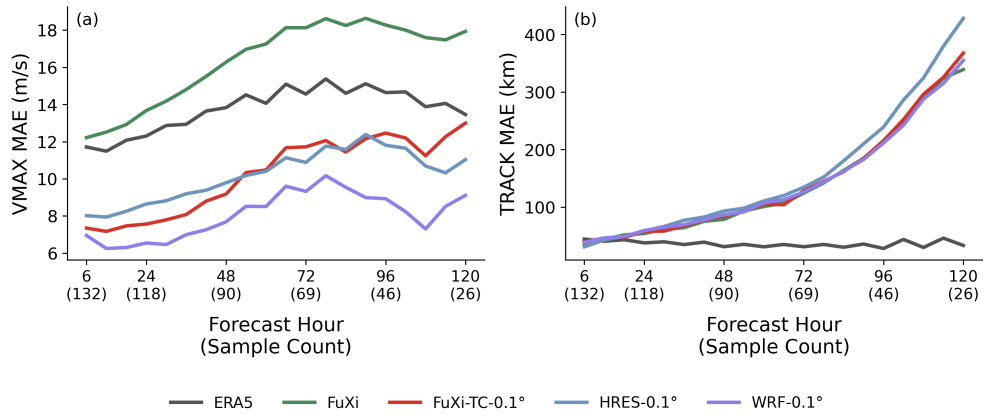


Fig. 1 Comparison of TC intensity and track forecasts with IBTrACS dataset as ground truth. Mean absolute error (MAE) of maximum sustained wind and tracks with forecast lead time for ERA5 (black lines), FuXi (green lines), FuXi-TC-0.1° (red lines), HRES-0.1° (blue lines) and WRF-0.1° (purple lines). All the forecast data are evaluated against the IBTrACS dataset. The evaluation covers forecasts for 17 TCs from July to October 2024.

Beyond track and intensity, accurate rainfall prediction is crucial for assessing the potential impact of a typhoon. Fig. 2 shows the time series of the Critical Success Index (CSI) of FuXi, FuXi-TC-0.1°, HRES-0.1° and WRF-0.1° for 6-hour accumulated total precipitation (TP) evaluated using various threshold values for 5-day forecasts. As expected, CSI values decrease with increasing thresholds, indicating greater challenges in predicting more extreme events. Compared to FuXi forecasts, the physics-based model (including HRES-0.1° and WRF-0.1°) maintain a distinct advantage in forecasting heavier precipitation events. When comparing CSI between FuXi and WRF-0.1°, WRF-0.1° significantly outperforms FuXi at higher thresholds, however FuXi achieves better CSI scores across a broader range of lower thresholds. The figure illustrates that FuXi-TC-0.1° successfully integrates the superior performance of physics-based models in capturing heavy precipitation with the proficiency of FuXi model in predicting general rainfall (indicated by the $\geq$ 0.1 mm threshold). Consequently, it achieves the highest CSI scores across most lead times and precipitation thresholds, suggesting its superior capabilities in precipitation forecasting. Notably, the CSI metrics exhibit a sawtooth-like fluctuation, which is particularly evident in the evaluation with a threshold greater than 0.1 mm. This oscillation stems from differences in valid data availability from AWS observations between 00/12 UTC and 06/18 UTC (see Supplementary Fig. 2), leading to significant fluctuations in forecast false alarms and corresponding variations in the probability of detection (Supplementary Figs. 5 and 6).

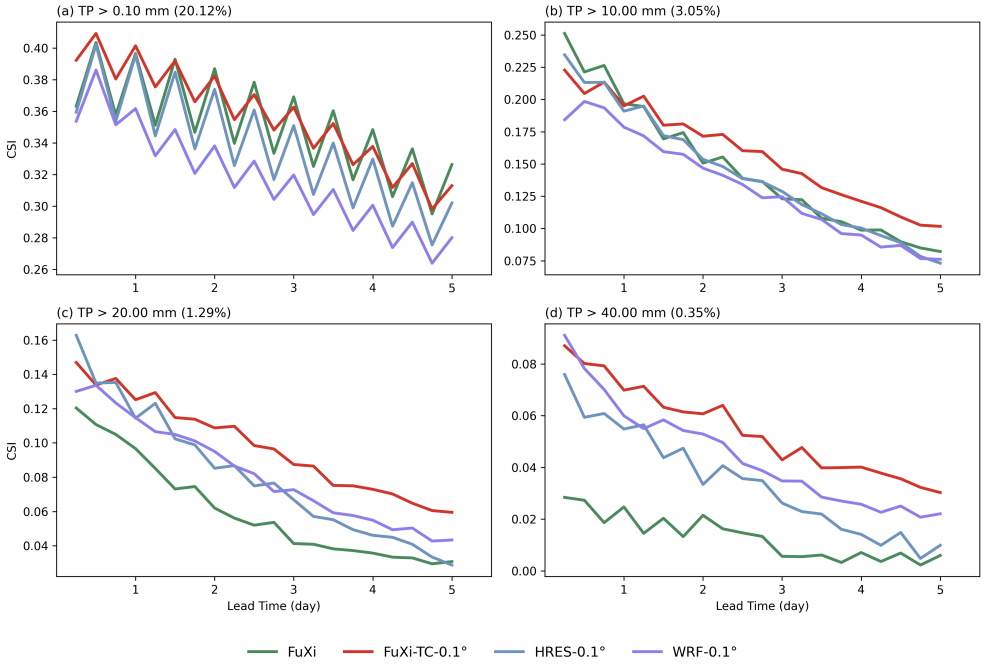


Fig. 2 Comparison of CSI for TP. Comparison for the FuXi (green lines), FuXi-TC-0.1° (red lines), HRES-0.1° (blue lines) and WRF-0.1° (purple lines) in predicting 6-hour precipitation for 0.1, 10, 20, 40 mm thresholds in (a-d) respectively. The ratios of extreme cases relative to the entire test set are indicated in parentheses. All the forecast data are evaluated against the Automatic Weather Station observations in period 1 July to 30 September 2024.

2.2 Deterministic case studies

To further illustrate the advantages of FuXi-TC-0.1°, we examine Typhoon Bebinca (2024254N10148) and Gaemi (2024201N12133) separately as representative cases of wind speed and precipitation intensity.

On 16 September 2024 at 07:30 Beijing time, Bebinca made landfall in Shanghai as a severe typhoon (42 m s^{-1}), marking the strongest typhoon to strike Shanghai and Jiangsu since 1949 and causing severe wind and rainfall impacts [32]. Fig. 3 shows the forecast initialized at 12:00 UTC on 15 September 2024. This specific time period was selected as an illustrative example because it covers not only the landfall of Typhoon Bebinca but also coincides with the peak intensity of its lifecycle. The figure shows that both ERA5 and FuXi dataset underestimate the intensity of Bebinca. FuXi-TC-0.1° presents stronger wind speeds like HRES-0.1° also contains finer-scale details in contrast to ERA5 and FuXi. The corresponding precipitation forecast for this case is provided in Supplementary Fig. 3, and the typhoon track forecast is shown in Supplementary Fig. 12.

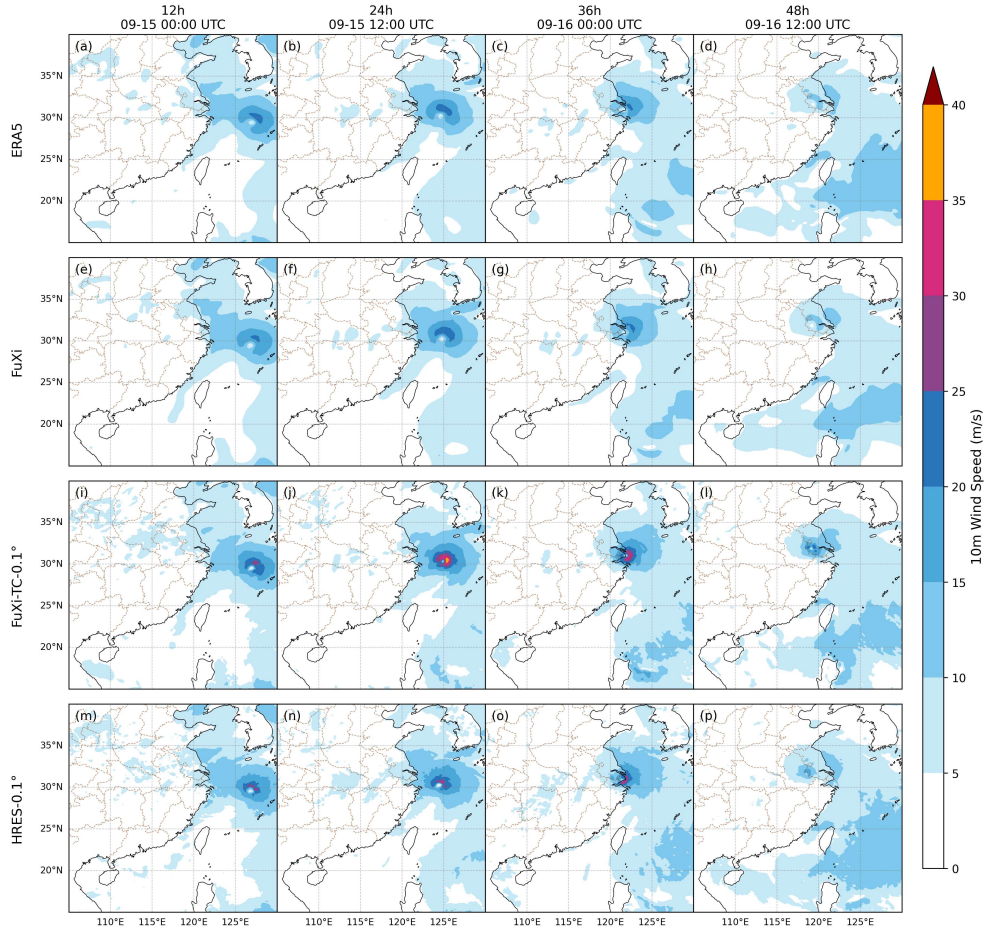


Fig. 3 Comparison of WS10M for Typhoon Bebinca forecasts. Comparison of snapshot examples of WS10M among ERA5 (first row), FuXi (second row), FuXi-TC-0.1° (third row) and HRES-0.1° (fourth row) for 12 (first column), 24 (second column), 36 (third column) and 48 (fourth row) hours forecasts with the initial time at 12:00 UTC 15 September 2024.

Fig. 4 displays the 24-hour accumulated precipitation observations from Automatic Weather Station (AWS) alongside forecasts initialized at 00 UTC on 23 July 2024 from FuXi, FuXi-TC-0.1°, WRF-0.1° and HRES-0.1°, over the periods of 24-48, 48-72, 72-96 and 96-120 hours. This case study was selected to exemplify the models' performance during Super Typhoon Gaemi, which triggered record-breaking extreme rainfall since 1961 as it made landfall in Fujian province [32]. For clarity, precipitation values are shown exclusively for mainland China, with oceanic and external landmasses masked. The figure reveals 0.1°-resolution models generate more detailed rainfall forecasts compared to FuXi (0.25°). Regarding the prediction of the heavy precipitation areas, FuXi-based forecasts align better with the observation, compared

to HRES-0.1°, especially at lead times of 4-5 days. In terms of rainfall intensity, FuXi evidently underestimates the peak values relative to observations. In contrast, FuXi-TC-0.1° yields a more accurate intensity forecast that agrees with observations and is comparable to WRF-0.1°. Furthermore, while WRF-0.1° produces more false alarms in many non-precipitation areas compared to FuXi, FuXi-TC-0.1° demonstrates higher accuracy in these rain-free regions. The corresponding track forecast for this case is provided in Supplementary Fig. 12.

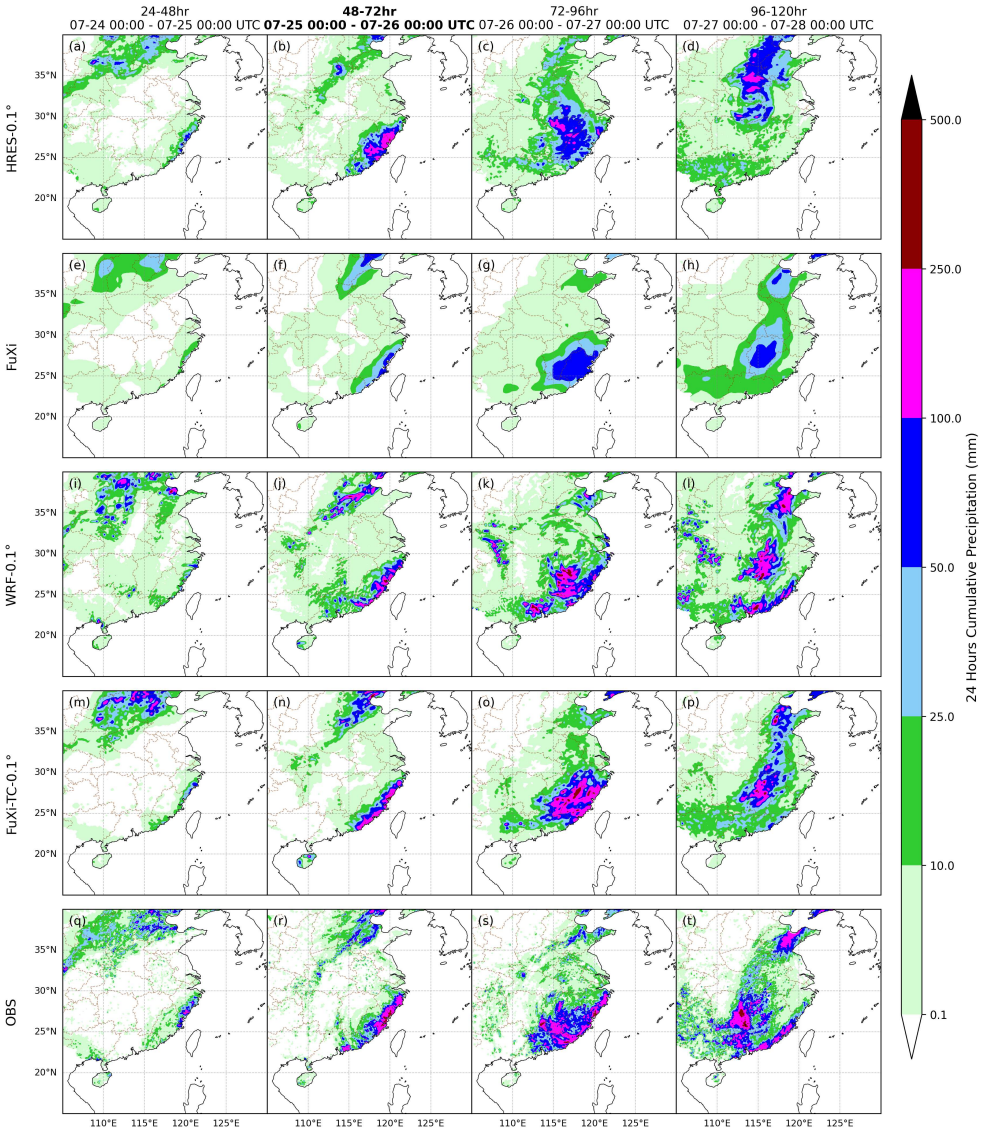


Fig. 4 Comparison of TP for Typhoon Gaemi. Comparison of snapshot examples of 24-hour accumulated TP among HRES-0.1° (first row), FuXi (second row), WRF-0.1° (third row), FuXi-TC-0.1° (fourth row) and the Automatic Weather Station observations (fifth row) mainland in China for 24-48 (first column), 48-72 (second column), 72-96 (third column) and 96-120 (fourth row) hours forecasts with the initial time at 00:00 UTC 23 July 2024.

2.3 Ensemble forecast comparison

Unlike deterministic forecasts, ensemble forecasts offer the distinct advantage of providing probabilistic guidance. While FuXi-ENS demonstrates superior performance over ECMWF-ENS with a comprehensive evaluation in key forecast metrics such as Continuous Ranked Probability Score (CRPS) and Brier Score (BS), challenges regarding ensemble under-dispersion and insufficient uncertainty quantification for extreme phenomena persist [31]. Following the FuXi-TC framework, a diffusion generative module is utilized to refine each FuXi-ENS member, yielding the enhanced ensemble system designated as FuXi-TC-ENS, which is then compared against ECMWF-ENS and FuXi-ENS in terms of TC intensity forecasts using 51 ensemble members. Fig. 5 illustrates temporal evolution of max sustained wind speed forecasts for Typhoon Bebinca. FuXi-ENS exhibits a notable underestimation of maximum sustained 10-m wind speeds compared to IBTrACS data. Specifically, the model's simulated peak wind speeds are limited to approximately 20 m/s, whereas observed values exceed 40 m/s. Furthermore, FuXi-ENS displays insufficient ensemble spread, which is likely due to the limited projection of initial perturbations and the model's inability to sufficiently amplify forecast perturbations over time [16]. Unlike simple additive noise, the diffusion-based refinement module enhances each ensemble forecast by superimposing generative perturbations onto FuXi-ENS. Consequently, FuXi-TC-ENS improves peak intensity estimation and exhibits a more realistic increase in ensemble spread with lead time without degrading track accuracy. The corresponding ensemble track forecasts for this case are provided in Supplementary Fig. 4.

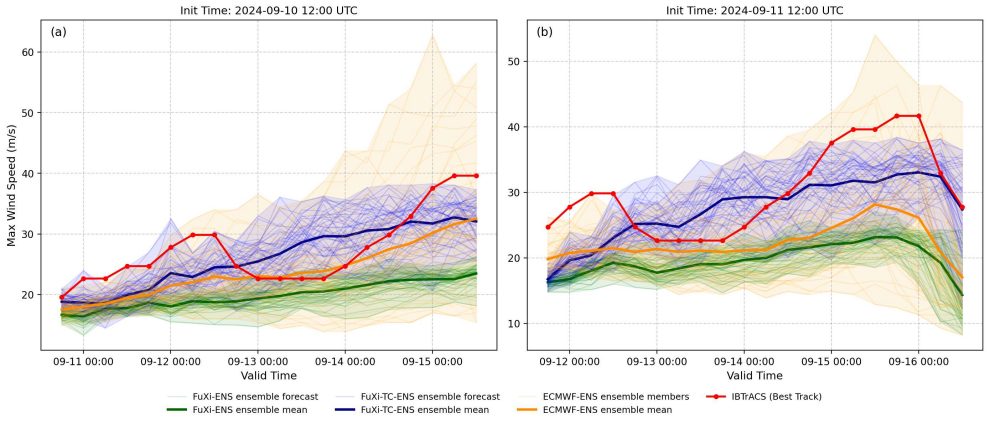


Fig. 5 Comparison of maximum sustained wind speed forecasts for Typhoon Bebinca. Comparison the maximum sustained wind speed from IBTrACS (red line) and ensemble forecasts generated by FuXi-ENS (green line), FuXi-TC-ENS (blue line) and ECMWF-ENS (orange line) with the initial forecasting time at 12:00 UTC 10 September and 12:00 UTC 11 September 2024.

2.4 Generalization to the NA Basin

To evaluate the generalization capability of FuXi-TC, we directly applied the model trained on the WNP to the NA region. Constrained by the fixed spatial dimensions of the WNP training domain, we partitioned the NA region into two overlapping subdomains: a western segment and an eastern segment (see Supplementary Fig. 11). We then reconstructed the full-field forecast for the entire NA basin by stitching these subdomains together, where predictions in the overlapping areas were averaged to ensure spatial continuity. By analyzing 15 TCs (see Supplementary Table 1) that occurred from June to November 2024, we present a statistical comparison of intensity and track forecasting performance among HRES-0.1°, ERA5, FuXi and FuXi-TC-0.1° in Fig. 6. In terms of TC intensity forecasting, FuXi-TC-0.1° still significantly outperforms FuXi and ERA5 in predicting the maximum WS10M. Moreover, FuXi-TC-0.1° delivers superior performance to HRES-0.1° within the first 36 hours and last 12 hours of forecasting, yet it underperforms HRES-0.1° in the mid-term forecasting period. Regarding TC track forecasting, compared with HRES-0.1°, the results of FuXi-based models (including FuXi and FuXi-TC-0.1°) exhibit better performance during the 48–120h forecasting window. To further illustrate this generalization capability, detailed case studies of North Atlantic hurricanes Beryl and Helene are provided (Supplementary Figs. 9 and 10). These cases demonstrate that FuXi-TC-0.1° successfully reconstructs intense, fine-scale inner-core wind structures, even in unseen geographical basins.

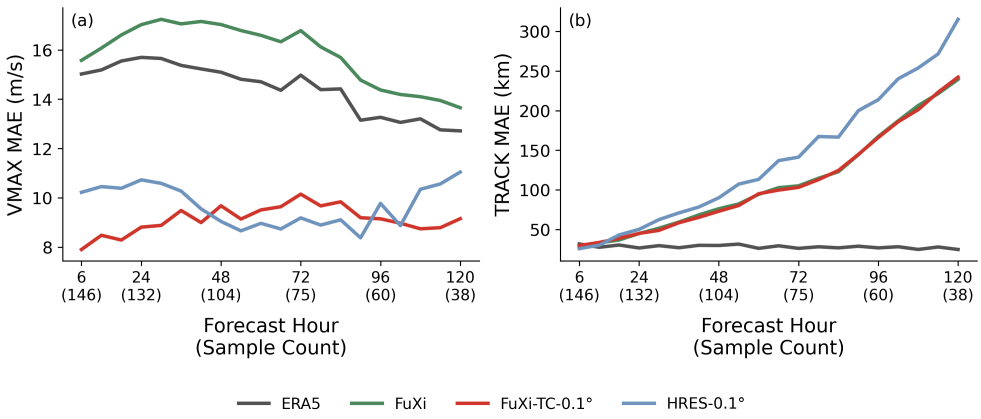


Fig. 6 Comparison of TC intensity and track forecasts in North Atlantic. MAE of maximum sustained wind and tracks with forecast lead time for ERA5 (black lines), FuXi (green lines), FuXi-TC-0.1° (red lines) and HRES-0.1° (blue lines). All the forecast data are evaluated against the IBTrACS dataset. The evaluation covers forecasts for 15 TCs from June to November 2024.

2.5 Computational efficiency comparison

Compared with traditional NWP models, a primary advantage of deep learning approaches is their drastically improved computational efficiency and reduced resource demands, particularly for large-domain and high-resolution forecasts. As detailed in Table 1, the WRF model at 0.1° resolution requires approximately 11.5 hours on 32 CPUs to complete a single 120-hour forecast under typical operational configurations. In contrast, FuXi-TC-0.1° generates a forecast with equivalent lead time and resolution in just 2 minutes on a single GPU, corresponding to a speedup of over two orders of magnitude. This efficiency enables rapid update cycles, facilitates large-scale ensemble forecasting, and significantly lowers the computational cost of high-resolution TC prediction, making FuXi-TC-0.1° feasible for real-time operational applications.

Table 1 Resource requirements and runtime per task.

Task	Resource	Quantity	Runtime (per forecast)
WRF-0.1°	Intel Xeon Platinum 8369B CPU @ 2.90GHz	32	11.5 hours
FuXi-TC-0.1°	NVIDIA A100 GPU	1	2 minutes

3 Discussion

Accurate prediction of TC intensity remains a significant challenge for both NWP and emerging AI forecasting systems. While recent AI–based weather forecasting models have achieved substantial improvements in track prediction, they often underestimate TC maximum wind speeds and associated precipitation. A major contributing factor is the intrinsic smoothing present in commonly used training datasets, particularly ERA5, which often suppresses extreme wind speeds and precipitation and blurs fine-scale structures associated with intense storms. In this study, we introduce FuXi-TC, a diffusion-based generative forecasting framework designed to mitigate these limitations. The model improves predictions of WS10M, TP and MSL while preserving the high accuracy of underlying FuXi-2.0 track forecasts. By recovering sharper spatial structures and strong extremes, the approach provides a computationally efficient alternative to traditional dynamical downscaling methods. Although WRF simulations are not bias-free, they provide dynamically consistent high-resolution structures that are absent in reanalysis datasets, making them suitable training targets for mesoscale refinement.

To address the smoothing bias prevalent in AI predictions, previous studies have primarily resorted to directly dynamical downscaling using NWP models. Despite their efficacy in recovering small-scale details, the prohibitive computational expense of such methods renders them difficult to scale for large-ensemble configurations or operational use. Beyond computational efficiency, a key advantage of the FuXi-TC framework is its robust transferability across both base forecasting models and geographical domains. Although trained exclusively on deterministic FuXi-2.0 forecasts

over the WNP, the generative module can be directly applied to other systems and regions. For instance, while the original FuXi-ENS achieves strong general forecast metrics, it suffers from significant under-dispersion and tends to underestimate the intensity of extreme events. Applying FuXi-TC to refine individual ensemble members of FuXi-ENS produces the FuXi-TC-ENS system, which increases ensemble spread without degrading track consistency and improves the representation of high-impact extremes. Unlike recent hybrid ensemble methods that rely on computationally expensive full physics-based models to couple with machine-learned flows [33], our framework achieves this enhancement without incurring massive computational costs. Its lightweight and highly transferable nature allows it to be seamlessly extended to large-ensemble scenarios.

Additionally, the model exhibits strong zero-shot geographical generalization. Despite being trained on the WNP basin, FuXi-TC performs robustly in the NA basin without fine-tuning, suggesting that the model may capture transferable physical structures associated with TC evolution rather than simply reproducing region-specific patterns. More broadly, this work highlights the potential of combining AI models with on NWP-derived data to extend the capabilities of AI weather forecasting. Recent analyses have suggested that the supply of high-quality human-generated training data for AI systems is becoming increasingly limited [34]. Synthetic data generated by models can partially address this constraint, but over-reliance on such data may degrade model performance and diminish creativity and introduce bias [35, 36]. In weather forecasting, the success of AI models has relied on extensive, high-quality, open-source datasets, such as ERA5, which span multiple decades. However, these datasets are finite, and the ERA5 archive has effectively been exhausted. Therefore, NWP models provide a complementary source of synthetic data that can augment training datasets and expose AI models to rare or extreme weather scenarios, e.g., TC [37]. Indeed, ERA5 itself can be viewed as a synthetic dataset, produced by ECMWF through 4D-Var data assimilation combined with CY41R2 model forecasts.

Beyond advancing TC forecasts, we demonstrate the broader significance of leveraging NWP-derived datasets to push the performance limits of AI-based weather forecast. However, a critical limitation of this framework is that FuXi-TC's performance upper bound may ultimately be locked by the systematic biases inherent in NWP physical parameterizations. To break through this theoretical ceiling, our future work will focus on directly integrating real-world observations to correct these NWP-induced biases.

4 Method

4.1 Dataset

4.1.1 Tropical cyclone dataset

To evaluate TC intensity and trajectory forecasts, we employ IBTrACS, a globally harmonized TC database maintained by NOAA National Centers for Environmental Information (NCEI), which integrates best-track records from multiple meteorological agencies [24, 38]. The dataset includes TC center position, maximum sustained surface

wind speed, and minimum central pressure (hPa). For validation against IBTrACS, we utilized the CMA agency best track data in the WNP and USA agency best track data in the NA respectively (see Supplementary Fig. 11 for the spatial domains).

To derive the forecasted TC tracks and intensities, we employ a modified version of the Aurora tracker [39]. The tracker identifies the TC center by detecting local minima in the MSL field—with 700 hPa geopotential height (Z700) as a fallback. Once the center is located, the forecasted intensity is determined by extracting the minimum MSL and maximum WS10M within a specific radius. Following the configuration of FuXi-Extreme [16], we incorporate physical termination criteria to prevent spurious tracking artifacts when the cyclone dissipates. Specifically, tracking is halted if the maximum 850 hPa relative vorticity and maximum WS10M falls below specific threshold values. This approach ensures alignment with standard evaluation protocols for TC forecasting.

For comparison with the ECMWF High-Resolution Forecast including both deterministic (HRES) and ensemble forecasts (ENS) efficiently, we acquired TC track data from the THORPEX Interactive Grand Global Ensemble (TIGGE) archive [40, 41]. This dataset incorporates tracks generated by the operational ECMWF tracker and is distributed in XML format, where the ECMWF HRES and ENS tracks are explicitly identified by the "Forecast" and "ensembleForecast" tags, respectively.

4.1.2 FuXi-2.0 dataset

Serving as the primary source of training dataset, FuXi-2.0 represents a significant advancement over its predecessor FuXi-1.0 [30], serving as an AI-based medium-range weather forecasting model with a coupled ocean–atmosphere structure. A key innovation is FuXi-2.0 employs a dual-model framework where a transformer-based interpolator efficiently derives continuous 1-hourly forecasts from 6-hourly predictions, thereby minimizing iterative steps and ensuring temporal coherence.

Furthermore, by explicitly modeling ocean feedback processes like sea surface heat fluxes, FuXi-2.0 improves the representation of air–sea interactions, which is critical for capturing the key features of TCs [42]. The model is trained and driven using the ERA5 reanalysis dataset, covering the period from 2012 to 2017.

This study uses FuXi-2.0 fields as the main source of global atmospheric initial and boundary conditions for the WRF model, with soil variables and sea surface temperature (SST) provided by ERA5 (A summary of variable definitions can be referred to in Table 2). The following surface variables are used: mean sea level pressure (MSL), 2-meter temperature (T2M), 2-meter dew point temperature (D2M), surface pressure (SP), and 10-m zonal and meridional winds (U10M and V10M). Upper-air variables on 13 standard pressure levels (50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, and 1000 hPa) include geopotential (Z), temperature (T), zonal wind (U), meridional wind (V), and relative humidity (RH), enabling a comprehensive representation of the tropospheric state. To ensure alignment with the variable requirements of the WRF model, RH is derived from specific humidity (Q), pressure (P), and temperature (T) using standard thermodynamic conversions. First, the actual vapor pressure

is calculated as follows [43]:

$$e = \frac{\mathrm{Q} \times \mathrm{P}}{0.622 + 0.378\mathrm{Q}} \tag{1}$$

where Q is the specific humidity in kg/kg, and P is the atmospheric pressure in hPa. Next, the saturation vapor pressure e_s (in hPa) is estimated using the empirical formula proposed by Bolton [44]:

$$e_s = 6.112 \exp\left(\frac{17.67\,\mathrm{T_c}}{\mathrm{T_c} + 243.5}\right) \tag{2}$$

where $\mathrm{T_c}$ is the temperature in °C ($\mathrm{T_c} = \mathrm{T} - 273.15$). Finally, the relative humidity is calculated as the ratio of actual to saturation vapor pressure, capped at 100%:

$$\mathrm{RH} = \frac{e}{e_s} \times 100\% \tag{3}$$

Table 2 Summary of initial and boundary condition variables for the WRF model.

Category	**Variables**	**Source**
single level	mean sea level pressure	FuXi-2.0
	surface pressure	FuXi-2.0
	2-m temperature	FuXi-2.0
	2-m dew point	FuXi-2.0
	10-m zonal/meridional winds	FuXi-2.0
	soil variables	ERA5
	sea surface temperature	ERA5
upper-air (13 pressure levels)[a]	geopotential	FuXi-2.0
	temperature	FuXi-2.0
	zonal/meridional wind	FuXi-2.0
	specific humidity	FuXi-2.0
	relative humidity	FuXi-2.0(Derived)

[a] 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, and 1000 hPa.

4.1.3 WRF dataset

High-resolution training targets are created using simulations from the WRF model (detailed parameter configurations are provided in Section 4.5). In contrast to previous hybrid approaches [13, 20, 27, 45] that initialize with Pangu-weather or Graphcast to provide initial and boundary conditions, the simulations in this study were driven primarily by FuXi-2.0 forecasts. This configuration preserves structural consistency between the large-scale input forecasts and the high-resolution target fields used for model training.

The WRF simulations were configured on a regular latitude-longitude grid with a horizontal resolution of 0.1°, comprising 592×496 grid points in longitude and latitude. In the vertical direction, the model atmosphere was discretized into 56 eta levels, extending to a model top pressure of 50 hPa. The downscaled fields provide spatially detailed representations of TC structure, including strong wind gradients and precipitation patterns that are not resolved in coarser and smoother AI forecasts. We select a set of key surface meteorological variables, including T2M, U10M, V10M, WS10M, MSL, and 6-hourly accumulated TP. Upper-air variables are extracted for pressure levels of 200, 300, 500, 700, and 850 hPa, including T, Q, U and V. Upper-air variables were vertically interpolated from model levels to the specified pressure levels using linear interpolation. These variables are integrated into the FuXi-TC framework to effectively capture the feature of TC.

4.1.4 Training data preprocessing

Preprocessing was applied separately to different variable types. For non-precipitation variables, Z-score normalization is applied using the mean and standard deviation computed from the training dataset. For TP, which exhibits a highly skewed, long-tailed distribution, we employ a logarithmic transformation to mitigate the dominance of extreme values and yield a more Gaussian-like distribution [46, 47]. Specifically, the transformation is defined as:

$$\tilde{x} = \ln(x + 1) \tag{4}$$

where x is the original 6-hourly accumulated precipitation, and $\tilde{x}$ is the transformed value. The constant 1 is added to handle zero precipitation values.

4.2 FuXi-TC model structure

As illustrated in Fig.7, the FuXi-TC framework employs a hybrid architecture that integrates global large-scale meteorological guidance with a generative refinement module. This design is inspired by the recent success of the FuXi-Extreme model [16], which demonstrated that coupling autoregressive forecasting with diffusion models significantly improves the representation of extreme weather events. Specifically, the framework extracts regional subsets from the FuXi-2.0 global forecast dataset (see regional domain in Supplementary Fig. 11). These coarse-resolution forecasts serve as the primary conditioning inputs for the downscaling process. In addition to these meteorological fields, the generative process is further conditioned on the diffusion noise level and various temporal features, including the forecast step, day of the year, and hour of the day, are encoded using sine/cosine transformations to accurately guide the denoising trajectory.

The final high-resolution forecasts are achieved via a Denoising Diffusion Probabilistic Model (DDPM) [28, 29], which is trained to map the coarse FuXi-2.0 priors to physically consistent WRF simulations.

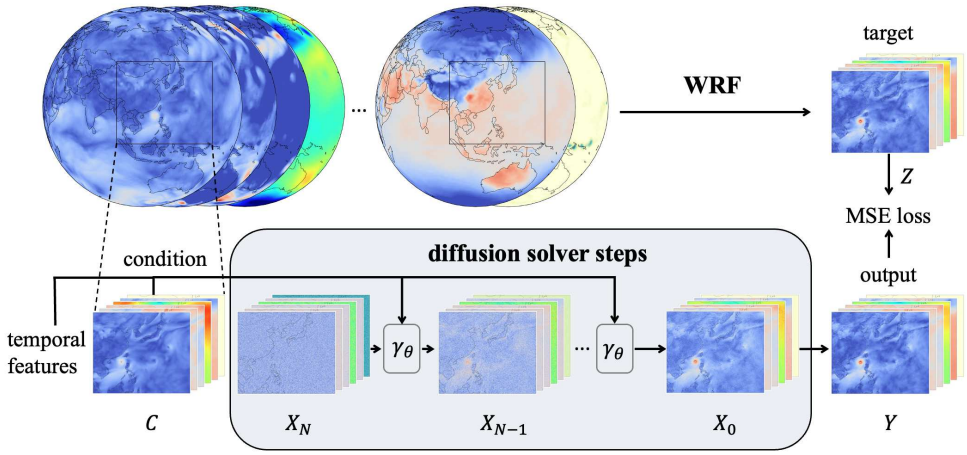


Fig. 7 Schematic diagram of the structures of the FuXi-TC framework. Coarse-resolution regional subsets are extracted from the global FuXi-2.0 forecasts to serve as the primary conditioning inputs. The generative refinement module, based on DDPM iteratively denoises random noise through a UNet-based solver. The denoising process is jointly conditioned on the meteorological fields and encoded temporal features. During training, the model is optimized using a MSE loss to map the generated high-resolution forecasts to physically consistent WRF simulations.

4.3 Denoising diffusion probabilistic model (DDPM)

The DDPM serves as the generative backbone of FuXi-TC, which treats the downscaling and bias-correction task as a conditional generation problem, similar to image super-resolution strategies employed in computer vision [48].

We employ the UNet [49] architecture from the Diffusers[1] library to build our DDPM model, which includes three downsampling layers and three upsampling layers. Each layer applies 2D convolutions with 3×3 kernels, SiLU activation functions [50] and group normalization[51], and the number of channels are 256, 512 and 1024 for the first, second, and third layers respectively. At the bottleneck of the network, 12 global self-attention layers [52] are integrated to enhance the capture of long-range spatial dependencies, which is crucial for restoring TC fine structures such as eye walls and spiral rain bands.

Formally, the diffusion process consists of a forward noise-injection process and a reverse denoising process. In the forward process, random Gaussian noise ϵ is gradually added to the ground-truth WRF data (Z) over t steps ($t \in [1, T]$) according to a variance schedule β_t [53]. The noisy sample X_t at step t is expressed as follows,

$$X_t = \sqrt{\bar{\alpha}_t} X_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \tag{5}$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t$ is the cumulative product of α, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, $\alpha_i = 1 - \beta_i$.

[1] https://github.com/huggingface/diffusers

In the reverse process, the model learns to predict the added noise: $\epsilon_\theta(X_t, t, C)$ with respect to the input condition C. The posterior distribution $\gamma_\theta(X_{t-1}|X_t, C)$ is assumed to be Gaussian, whose mean μ parameterized by the UNet model with parameters θ:

$$\mu_\theta(X_t, t, C) = \frac{1}{\sqrt{\alpha_t}} \left(X_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(X_t, t, C) \right) \tag{6}$$

By predicting the noise, the model learn to reconstruct the original data from the sampled noise X_T via iterative denoising.

The loss function of the DDPM model adopts mean squared error (MSE), which minimizes the difference between the noise added in the forward process and the noise predicted by the model.

$$L = \mathbb{E}_{t, X_0, \epsilon, C} \left[||\epsilon - \epsilon_\theta(X_t, t, C)||^2 \right] \tag{7}$$

During inference, sampling was performed using 5 denoising steps with a linear beta schedule. This configuration provides a balance between computational efficiency and reconstruction fidelity.

4.4 FuXi-TC model training

The FuXi-TC model is trained based on the PyTorch [54] framework, and the training process only updates the parameters of the DDPM model, while the parameters of FuXi remain frozen to ensure the stability of the base forecast. The training is carried out on 2 Nvidia A100 GPUs, with a batch size of 2 per GPU and a total of 60000 iterations. We use the AdamW [55] optimizer for training, with the parameters set as $\beta_1 = 0.9$, $\beta_2 = 0.95$. We select the initial learning rate of $2.5{\times}10^{-4}$ to balance the convergence speed and stability, and use the weight decay coefficient 0.1 to suppress overfitting.

4.5 WRF model configuration

We employ WRF version 4.3 [56] to conduct regional simulations over the WNP (see Supplementary Fig. 11). The model is initialized twice daily at 00 and 12 UTC, with forecasts extending up to 120 hours. To facilitate numerical integration and subsequent post-processing training framework, a latitude–longitude projection is adopted. The simulation domain spans 100°E–159.1°E and 0.5°N–50°N, a coverage sufficient to capture the majority of TC tracks in the WNP. This corresponds to a horizontal grid dimension of 592 × 496 points with 0.1° horizontal interval. In the vertical direction, the model top is set at 50 hPa, with 56 vertical levels defined by eta coordinates.

We evaluated the model's sensitivity to horizontal resolution by comparing simulations at 0.25° and 0.1° using identical physical configurations. Supplementary Fig. 1 illustrates the comparative performance regarding TC track and intensity forecasting. The results indicate that while track forecast accuracy remains similar between the two resolutions, the WRF-0.1° configuration demonstrates significantly superior performance in predicting TC intensity compared to WRF-0.25°. Consequently, the 0.1° horizontal resolution was selected for this study. The WRF outputs at this selected

resolution are generated for specific periods in 2023 and 2024. The WRF outputs at this selected resolution are generated for specific periods. We utilize data from April 1 to October 31 of 2023 for model training and reserve data from July 1 to October 31 of 2024 for independent testing.

Both physical parameterization and spectral nudging schemes follow those used in the Shanghai Typhoon Model (SHTM) [57]. Specifically, the model employs the Thompson microphysics scheme [58], the multi-scale Kain–Fritsch cumulus scheme [59], the RRTMG scheme for longwave radiation [60], the unified Noah land-surface model [61], and the Yonsei University planetary boundary layer scheme [62].

Spectral nudging is applied to maintain consistency between the WRF simulations and FuXi's large-scale circulation while allowing the model to develop its own mesoscale structures. The nudging configuration is optimized in terms of nudged variables, vertical application, horizontal wavelength cutoff, and relaxation time. Only the zonal and meridional wind components and virtual temperature are nudged, while specific humidity is excluded to prevent degradation of TC intensity forecasts [27].

Vertically, nudging is applied only above 850 hPa, excluding the planetary boundary layer to avoid adverse effects arising from FuXi's limited vertical detail near the surface and from differences in surface forcing over complex terrain. A height-dependent weighting reduces the nudging influence toward the surface, thereby preserving boundary-layer processes[27].

In spectral space, only large-scale features exceeding a specified wavelength threshold are constrained, ensuring that synoptic-scale patterns from FuXi are retained while WRF generates smaller-scale variability[57]. A relatively weak relaxation coefficient provides gentle constraints, enabling the model to align with FuXi's large-scale fields without suppressing mesoscale dynamics. Nudging is activated from the start of the integration and maintained throughout the forecast period to ensure persistent large-scale alignment.

4.6 Evaluation method

4.6.1 TC tracking method

In our study, we adopt a modified version of Aurora tracker [39] to all forecast model's performance including FuXi, FuXi-TC, FuXi-ENS and FuXi-TC-ENS. Initially, we outline the Aurora tracker, then detail our modifications and the justification for these changes.

The primary function of the Aurora tracker is to identifies TC center positions by detecting local minima in predicted MSL fields, with Z700 employed as a fallback variable when MSL-based detection fails when plagued by multiple local minima. The tracker is initialised at the observed TC position from IBTrACS best track data (China Meteorological Administration agency for WNP and United States agency for NA) at the analysis time. For each subsequent 6-hourly forecast lead time, a first-guess position is estimated by linear extrapolation of the preceding tracked positions (up to the most recent eight estimates, corresponding to 48 h). To locate the TC center, the tracker searches for the local minimum of the Gaussian-smoothed MSL field closest to the first-guess position. Within a search box of $x^\circ \times x^\circ$ centered on the first-guess

position. The tracker sequentially attempts search boxes of 4°, 3°, 2°, and 1.5° radius until a valid interior local minimum is identified. If all MSL-based attempts fail, the procedure falls back to detecting the closest local minimum of Z700, and then repeats the procedure within varying box sizes for MSL.

Following the ECMWF's TC tracker algorithm described by Van Der Grijn [63], FuXi-Extreme tracking method [16], and informed by our analysis of mistracks within our dataset, we introduce several enhancements to this algorithm as described below.

(1)In addition to utilizing the local minima of MSL and Z700, we incorporate two physical termination criteria based on the local maxima of WS10M and vorticity at 850 hPa to enhance the robustness of TC identification. Specifically, tracking is terminated if the maximum 850 hPa relative vorticity within a 1.5° range of the center falls below 5×10^{-5} s^{-1} or if the maximum WS10M drops below 8 m/s within the same range. This thresholding prevents spurious tracking artifacts that may arise when the cyclone's intensity dissipates.

(2)The models evaluated in our study operate at varying spatial resolutions, specifically 0.25° and 0.1°. To account for these differences, we linearly scale the parameters of the Gaussian low-pass filter according to the specific grid spacing of each dataset. This adjustment standardizes the effective smoothing area across all forecasts. Consequently, the physical scale for local minimum detection remains approximately $2° \times 2$ regardless of the native model resolution, maintaining consistent tracking behavior across forecast systems with different grid spacings.

4.6.2 TC track and intensity evaluation

Track accuracy is quantified using the MAE, which calculates distance between observed and predicted TC center positions. Intensity evaluation involves assessing the maximum sustained WS10M near the TC center using MAE. Phases categorized as Tropical Depression (TD) or those flagged as Extratropical Cyclone (ET) stages were explicitly excluded from the dataset to minimize tracking uncertainties associated with weak or asymmetric baroclinic systems. Consistent with the WNP methodology, the evaluation included valid time steps categorized as Tropical Storm, Hurricane, Tropical Cyclone, Typhoon, or Super Typhoon. This evaluation encompasses totally 35 TCs from 2024 in both WNP and NA only if the IBTrACS, European Centre for Medium-Range Weather Forecasts Reanalysis v5 (ERA5) and forecast results are simultaneously recorded.

4.6.3 Total precipitation evaluation

To quantitatively assess the forecast model performance, we utilize standard categorical verification statistics based on contingency table including True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN) [64]. The Critical Success Index (CSI), also known as the Threat Score (TS), provides a comprehensive measure of accuracy, accounting for both misses and false alarms. It is calculated using the equation $\mathrm{CSI} = \frac{\mathrm{TP}}{\mathrm{TP+FP+FN}}$. Probability of Detection (POD) score measures the fraction of observed events that were correctly predicted, using the equation

POD $= \frac{\mathrm{TP}}{\mathrm{TP+FN}}$. And False Alarm Ratio (FAR) score highlights the tendency of the dataset to overestimate precipitation occurrence, using the equation FAR $= \frac{\mathrm{FP}}{\mathrm{TP+FP}}$.

The 6-hour accumulated ground truth precipitation data on land were obtained from the AWS provided by the China Meteorological Data Service Centre (CMDC). Since the AWS network is limited to mainland China, a spatial masking procedure was applied to the forecast data. Therefore, grid points corresponding to oceanic regions and areas outside mainland Chinese territory were excluded to ensure spatial consistency.

Furthermore, direct comparison between point-based gauge observations and grid-based NWP outputs is prone to representativeness error due to scale mismatches [65]. To mitigate this and facilitate a rigorous grid-to-grid verification, we reference the standard practice to map observational data onto the model grid [66, 67]. Therefore, consistent with established verification frameworks, the valid station observations were interpolated to the 0.1° model's grid resolution prior to statistical calculation. The metrics were computed for precipitation accumulated over a 6-hour period, using threshold values of 0.1, 10, 20, and 40 mm to evaluate the model's performance across light to heavy rainfall intensities.

Data Availability Statement

The ERA5 reanalysis data are accessible through the Copernicus Climate Data Store at https://cds.climate.copernicus.eu/. ECMWF HRES and ENS TC track forecasts can be retrieved from https://confluence.ecmwf.int/display/TIGGE. We downloaded the ground truth tracks of TC from the International Best Track Archive for Climate Stewardship (IBTrACS) project, which is publicly available at https://www.ncei.noaa.gov/products/international-best-track-archive.

Code Availability Statement

The code base of the DDPM model used in this work is available at https://github.com/luchengthu/dpm-solver [68]. The source code used for running FuXi-TC model, along with the scripts for running the WRF simulations in this work, are available from the Zenodo link (https://zenodo.org/records/19382719 [69]). The source code used for training and running FuXi models are available at https://doi.org/10.5281/zenodo.8100201 [30]. Additionally, the basic Aurora tracker algorithm in this work can be downloaded from https://github.com/microsoft/aurora [39]. Furthermore, the source code for the WRF model version 4.3 is publicly available at https://github.com/wrf-model/WRF [56].

Acknowledgments

We gratefully acknowledge the ECMWF for their efforts in producing, maintaining, and distributing the ERA5 reanalysis and ECMWF HRES datasets. We acknowledged the efforts of NOAA National Centers for Environmental Information in making the IBTrACS dataset available. We recognize the efforts of CMA for sharing the IBTrACS

dataset. We thank support from the Computing for the Future at Fudan (CFFF), which has provided us with a high-performance computing platform. This work was supported by AI for Science Program, Shanghai Municipal Commission of Economy and Informatization, 2025-GZL-RGZN-BTBX-02017.

Competing interests

The authors declare no competing interests.

References

[1] Emanuel, K.: Increasing destructiveness of tropical cyclones over the past 30 years. Nature **436**(7051), 686–688 (2005) https://doi.org/10.1038/nature03906

[2] Emanuel, K., Chonabayashi, S., Bakkensen, L., Mendelsohn, R.: The impact of climate change on global tropical cyclone damage. Nature Climate Change **2** (2012) https://doi.org/10.1038/nclimate1357

[3] Zhang, Q., Wu, L., Liu, Q.: Tropical cyclone damages in china 1983–2006. Bulletin of the American Meteorological Society **90**(4), 489–496 (2009) https://doi.org/10.1175/2008BAMS2631.1

[4] Zhang, Q., Gu, X., Shi, P., Singh, V.P.: Impact of tropical cyclones on flood risk in southeastern china: Spatial patterns, causes and implications. Global and Planetary Change **150**, 81–93 (2017) https://doi.org/10.1016/j.gloplacha.2017.02.004

[5] Peduzzi, P., Chatenoux, B., Dao, H., De Bono, A., Herold, C., Kossin, J., Mouton, F., Nordbeck, O.: Global trends in tropical cyclone risk. Nature Climate Change **2**(4), 289–294 (2012) https://doi.org/10.1038/nclimate1410

[6] Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (eds.): Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, ??? (2021). https://doi.org/10.1017/9781009157896 . https://www.ipcc.ch/report/ar6/wg1/

[7] Webster, P., Holland, G., Curry, J., Chang, H.-R.: Changes in tropical cyclone number, duration, and intensity in a warming environment. Science **309**, 1844–1846 (2005) https://doi.org/10.1126/science.1116448

[8] Tallapragada, V.: In: Mohanty, U.C., Gopalakrishnan, S.G. (eds.) Overview of the NOAA/NCEP Operational Hurricane Weather Research and Forecast (HWRF) Modelling System, pp. 51–106. Springer, ??? (2016). https://doi.org/10.5822/978-94-024-0896-6_3 . https://doi.org/10.5822/978-94-024-0896-6_3

[9] Alaka, G.J., Sippel, J.A., Zhang, Z., Kim, H.-S., Marks, F.D., Tallapragada, V., Mehra, A., Zhang, X., Poyer, A., Gopalakrishnan, S.G.: Lifetime performance of the operational hurricane weather research and forecasting model (hwrf) for north atlantic tropical cyclones. Bulletin of the American Meteorological Society **105**(6), 932–961 (2024) https://doi.org/10.1175/BAMS-D-23-0139.1

[10] Wang, W., Han, J., Yang, F., Steffen, J., Liu, B., Zhang, Z., Mehra, A., Tallapragada, V.: Improving the intensity forecast of tropical cyclones in the hurricane analysis and forecast system. Weather and Forecasting **38**(10), 2057–2075 (2023) https://doi.org/10.1175/WAF-D-23-0041.1

[11] Ma, S., Zhang, J., Qu, A., Wang, D., Shen, X.: Impacts to tropical cyclone prediction of grapes_tym from increasing of model vertical levels and enlargement of model forecast domain. Acta Meteorologica Sinica **79**(1), 94–103 (2021) https://doi.org/10.11676/qxxb2020.067

[12] Cangialosi, J.P., Blake, E., DeMaria, M., Penny, A., Latto, A., Rappaport, E., Tallapragada, V.: Recent progress in tropical cyclone intensity forecasting at the national hurricane center. Weather and Forecasting **35**(5), 1913–1922 (2020) https://doi.org/10.1175/WAF-D-20-0059.1

[13] Xu, H., Zhao, Y., Zhao, D., Duan, Y., Xu, X.: Exploring the typhoon intensity forecasting through integrating ai weather forecasting with regional numerical weather model. npj Climate and Atmospheric Science **8**(1), 38 (2025)

[14] Xinyuan, B., Jinping, L., Yihong, D.: Review of artificial intelligence application in typhoon forecasting. Tropical Cyclone Research and Review (2025)

[15] Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., Tian, Q.: Accurate medium-range global weather forecasting with 3d neural networks. Nature **619**(7970), 533–538 (2023)

[16] Zhong, X., Chen, L., Liu, J., Lin, C., Qi, Y., Li, H.: Fuxi-extreme: Improving extreme rainfall and wind forecasts with diffusion model. Science China Earth Sciences, 1–13 (2024)

[17] Mahesh, A., Cohen, Y., Brenowitz, N., Elms, J., Subramanian, S., Harrington, P., Anandkumar, A., Pathak, J., Kurth, T., Bonev, B., *et al.*: Evaluating data-driven forecasts of extreme weather. In: AGU Fall Meeting Abstracts, vol. 2023, pp. 31–2513 (2023)

[18] Wang, J., Tabas, S., Yang, F., Levit, J., Stajner, I., Montuoro, R., Tallapragada, V., Gross, B.: Machine learning weather prediction model development for global ensemble forecasts at ncep. In: EGU General Assembly Conference Abstracts, p. 11707 (2024)

[19] Shi, Y., Hu, R., Wu, N., Zhang, H., Liu, X., Zeng, Z., Zhu, J., Han, P., Luo, C.,

Zhang, H., *et al.*: Comparison of ai and nwp models in operational severe weather forecasting: A study on tropical cyclone predictions. Journal of Geophysical Research: Machine Learning and Computation **2**(2), 2024–000481 (2025)

[20] Liu, H.-Y., Tan, Z.-M., Wang, Y., Tang, J., Satoh, M., Lei, L., Gu, J.-F., Zhang, Y., Nie, G.-Z., Chen, Q.-Z.: A hybrid machine learning/physics-based modeling framework for 2-week extended prediction of tropical cyclones. Journal of Geophysical Research: Machine Learning and Computation **1**(3), 2024–000207 (2024)

[21] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., *et al.*: The era5 global reanalysis. Quarterly journal of the royal meteorological society **146**(730), 1999–2049 (2020)

[22] Hodges, K., Cobb, A., Vidale, P.L.: How well are tropical cyclones represented in reanalysis datasets? Journal of Climate **30**(14), 5243–5264 (2017) https://doi.org/10.1175/JCLI-D-16-0557.1

[23] Zhang, X., Zuo, C., Wang, Z., Tao, C., Han, Y., Zuo, J.: Typhoon storm surge simulation study based on reconstructed era5 wind fields—a case study of typhoon "muifa", the 12th typhoon of 2022. Journal of Marine Science and Engineering **12**(11), 2099 (2024)

[24] Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J., Neumann, C.J.: The international best track archive for climate stewardship (ibtracs) unifying tropical cyclone data. Bulletin of the American Meteorological Society **91**(3), 363–376 (2010)

[25] Wang, X., Liu, L., Chen, K., Han, T., Li, B., Bai, L.: Vqlti: Long-term tropical cyclone intensity forecasting with physical constraints. arXiv preprint arXiv:2501.18122 (2025)

[26] Huang, C., Mu, P., Zhang, J., Chan, S., Zhang, S., Yan, H., Chen, S., Bai, C.: Benchmark dataset and deep learning method for global tropical cyclone forecasting. Nature Communications **16**(1), 5923 (2025)

[27] Husain, S.Z., Separovic, L., Caron, J.-F., Aider, R., Buehner, M., Chamberland, S., Lapalme, E., McTaggart-Cowan, R., Subich, C., Vaillancourt, P.A., et al.: Leveraging data-driven weather models for improving numerical weather prediction skill through large-scale spectral nudging. Weather and Forecasting **1**(aop) (2025)

[28] Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning, pp. 2256–2265 (2015). PMLR

[29] Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851 (2020)

[30] Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., Li, H.: Fuxi: a cascade machine learning forecasting system for 15-day global weather forecast. npj climate and atmospheric science **6**(1), 190 (2023)

[31] Zhong, X., Chen, L., Li, H., Liu, J., Fan, X., Feng, J., Dai, K., Luo, J.-J., Wu, J., Lu, B.: Fuxi-ens: A machine learning model for medium-range ensemble weather forecasting. arXiv preprint arXiv:2405.05925 (2024)

[32] Wu, L., Yu, R., Xiang, C., Yu, H., Feng, Y., Zhou, X.: Extreme Impacts of Four Landfalling Tropical Cyclones in China during the 2024 Peak Season. Springer (2025)

[33] Polichtchouk, I., Lang, S., Lock, S.-J., Maier-Gerber, M., Dueben, P.: Hybrid ensemble forecasting combining physics-based and machine-learning predictions through spectral nudging (2026). https://arxiv.org/abs/2603.05570

[34] Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., Hobbhahn, M.: Position: Will we run out of data? limits of LLM scaling based on human-generated data. In: Forty-first International Conference on Machine Learning (2024). https://openreview.net/forum?id=ViZcgDQjyG

[35] Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., Gal, Y.: Ai models collapse when trained on recursively generated data. Nature **631**(8022), 755–759 (2024) https://doi.org/10.1038/s41586-024-07566-y

[36] Gerstgrasser, M., Schaeffer, R., Dey, A., Rafailov, R., Sleight, H., Hughes, J., Korbak, T., Agrawal, R., Pai, D., Gromov, A., et al.: Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. Preprint at https://arxiv.org/abs/2404.01413 (2024)

[37] Bauer, P.: What if? numerical weather prediction at the crossroads. Journal of the European Meteorological Society **1**, 100002 (2024)

[38] Kenneth, R., Howard, J., James, P., Michael, C., Carl, J.: International best track archive for climate stewardship (ibtracs) project, version 4. (No Title) (2019)

[39] Bodnar, C., Bruinsma, W.P., Lucic, A., Stanley, M., Allen, A., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J.A., Dong, H., *et al.*: A foundation model for the earth system. Nature **641**(8065), 1180–1187 (2025)

[40] Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D.H., Ebert, B., Fuentes, M., Hamill, T.M., Mylne, K., *et al.*: The thorpex interactive grand global ensemble. Bulletin of the American Meteorological Society **91**(8), 1059–1072 (2010)

[41] Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T.M., Hewson, T.D., Keller, J.H., Matsueda, M., Methven, J., Pappenberger, F., *et al.*: The tigge project and its achievements. Bulletin of the American Meteorological Society **97**(1), 49–67 (2016)

[42] Zhong, X., Chen, L., Fan, X., Qian, W., Liu, J., Li, H.: Fuxi-2.0: Advancing machine learning weather forecasting model for practical applications. arXiv preprint arXiv:2409.07188 (2024)

[43] Wallace, J.M., Hobbs, P.V.: Atmospheric Science: an Introductory Survey, 2nd edn. Elsevier Academic Press, ??? (2006)

[44] Bolton, D.: The computation of equivalent potential temperature. Monthly weather review **108**(7), 1046–1053 (1980)

[45] Niu, Z., Huang, W., Zhang, L., Deng, L., Wang, H., Yang, Y., Wang, D., Li, H.: Improving typhoon predictions by integrating data-driven machine learning model with physics model based on the spectral nudging and data assimilation. Earth and Space Science **12**(2), 2024–003952 (2025)

[46] Kuhn, M., Johnson, K.: Applied Predictive Modeling vol. 26. Springer, ??? (2013)

[47] Shrestha, A., *et al.*: Evaluation of machine learning models for the routing of ensemble precipitation forecasts. Journal of Hydrology **570**, 646–657 (2019)

[48] Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(4), 4713–4726 (2022)

[49] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241. Springer, Cham (2015)

[50] Ramachandran, P., Zoph, B., Le, Q.V.: Searching for activation functions. In: International Conference on Learning Representations (2018)

[51] Wu, Y., He, K.: Group normalization. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)

[52] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)

[53] Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning, pp. 8162–8171 (2021). PMLR

[54] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen,

T., Lin, Z., Gimelshein, N., Antiga, L., *et al.*: Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, vol. 32 (2019)

[55] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)

[56] Jensen, A.A., Gill, D.O., Powers, J.G., Duda, M.G.: A description of the advanced research wrf model version 4.3. A Description of the Advanced Research WRF Model Version 4.3 (2021) NCAR/TN-556+ STR **556** (2021)

[57] Niu, Z., Wang, D., Mu, M., Huang, W., Fan, X., Yang, M., Qin, B.: Machine-learning (ml)-physics fusion model accelerates the paradigm shift in typhoon forecasting with a cnop-based assimilation framework. Geophysical Research Letters **52**(15), 2025–115926 (2025)

[58] Thompson, G., Rasmussen, R.M., Manning, K.: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. part i: Description and sensitivity analysis. Monthly Weather Review **132**(2), 519–542 (2004)

[59] Zheng, Y., Alapaty, K., Herwehe, J.A., Del Genio, A.D., Niyogi, D.: Improving high-resolution weather forecasts using the weather research and forecasting (wrf) model with an updated kain–fritsch scheme. Monthly Weather Review **144**(3), 833–860 (2016)

[60] Iacono, M.J., Delamere, J.S., Mlawer, E.J., Shephard, M.W., Clough, S.A., Collins, W.D.: Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models. Journal of Geophysical Research: Atmospheres **113**(D13) (2008)

[61] Ek, M., Mitchell, K., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J.: Implementation of noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model. Journal of Geophysical Research: Atmospheres **108**(D22) (2003)

[62] Hu, X.-M., Klein, P.M., Xue, M.: Evaluation of the updated ysu planetary boundary layer scheme within wrf for wind resource and air quality assessments. Journal of Geophysical Research: Atmospheres **118**(18), 10–490 (2013)

[63] Grijn, V.: Tropical cyclone forecasting at ecmwf: New products and validation. ECMWF Tech. Memo. **386**, 1 (2002)

[64] Wilks, D.S.: Statistical Methods in the Atmospheric Sciences, 3rd edn. Academic Press, Oxford (2011)

[65] Tustison, B., Harris, D., Foufoula-Georgiou, E.: Scale issues in verification of precipitation forecasts. Journal of Geophysical Research: Atmospheres **106**(D11),

11775–11784 (2001) https://doi.org/10.1029/2001JD900066

[66] Accadia, C., Mariani, S., Casaioli, M., Lavagnini, A., Speranza, A.: Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. Weather and forecasting **18**(5), 918–932 (2003)

[67] Accadia, C., Mariani, S., Casaioli, M., Lavagnini, A., Speranza, A.: Verification of precipitation forecasts from two limited-area models over italy and comparison with ecmwf forecasts using a resampling technique. Weather and Forecasting **20**(3), 276–300 (2005)

[68] Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., Zhu, J.: Dpm-solver++: Fast high-order solver for diffusion probabilistic models. arXiv preprint arXiv:2211.01095 (2022)

[69] Guo, S.: FuXi-TC. Zenodo (2026). https://doi.org/10.5281/zenodo.19382719 . https://doi.org/10.5281/zenodo.19382719

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFuXiTC.pdf.pdf
- nreditorialpolicychecklist.pdf