International Journal of Information Technology and Computer Science Applications (IJITCSA)
p-ISSN: 2964-3139 e-ISSN: 2985-5330
Vol. 04, No. 02, page 154 - 174
Submitted 16/4/2026; Accepted 26/7/2026; Published 27/7/2026

Physics-Aware AI-Initialized Dynamical Downscaling for Regional Extreme-Weather Forecasting with Open Climate Data

Mina Annetta[1*], Shanti Purohit[1], Viljar Vagle[1], Deepak Deo[1]
[1]Kalinga Institute of Industrial Technology, Patia, Bhubaneswar, Odisha, INDIA.
e-mail: minakiit9912@gmail.com

Corresponding Autor: Mina Annetta

Abstract

Accurate regional forecasting of extreme precipitation remains difficult because the scales that control disaster-producing rainfall are neither fully resolved by global numerical weather prediction nor reliably preserved by current global artificial intelligence weather models. Global AI systems such as Pangu-Weather, GraphCast, GenCast, and related models have transformed medium-range forecasting skill and computational efficiency, yet they remain fundamentally constrained by coarse training targets, regression-induced smoothing, and limited direct representation of terrain-locked convection and local hydrometeorological extremes. This paper reconstructs and substantially extends an event-based manuscript on AI-driven regional forecasting into a submission-oriented framework centered on the more defensible idea of physics-aware regional extreme-weather forecasting or downscaling with open climate data. The core argument is that AI should not be treated as a wholesale substitute for high-resolution regional physics; rather, it should be used as a skillful large-scale predictor whose state can be physically harmonized and injected into a regional nonhydrostatic model. We therefore formalize an AI-initialized, physics-aware dynamical downscaling pipeline in which open global reanalysis and observation products are used to generate, constrain, and evaluate regional forecasts of extreme rainfall. The framework is instantiated using the published North China July–August 2023 extreme precipitation case, for which the original study compared WRF simulations driven by Pangu forecasts against WRF simulations driven by NCEP GFS forecasts across lead times of 0.5, 3.0, and 5.5 days. This paper contributes in three ways. First, it repositions the original study within the modern literature on AI weather forecasting, regional downscaling, and physically constrained machine learning. Second, it formulates the coupling problem mathematically, clarifies the state alignment needed to make AI forecasts dynamically usable by WRF, and introduces a coherent reliability-oriented evaluation logic based on error growth, threshold skill, and event-structure consistency. Third, it reorganizes the experiments and results into a rigorous narrative grounded in reproducibility. Using the published event-level metrics, the AI-initialized regional system outperforms the GFS-initialized counterpart at extended lead times. For the North China case, the maximum precipitation threshold retaining a Threat Score of at least 0.1 is 400 mm at 5.5-day lead for Pangu-initialized WRF, whereas the GFS-driven counterpart retains comparable skill only at 50 mm. At 0.5-day lead, both systems perform competitively, but the AI-driven system still exhibits stronger spatial correlation (0.76 versus 0.68) and lower RMSE (86.2 mm versus 96.4 mm). The evidence supports a restrained but important conclusion: physics-aware AI-initialized regional modeling is a promising route for long-lead extreme-weather forecasting, yet current evidence remains case-limited and should be interpreted as a strong event-based demonstration rather than universal proof of general superiority.

Keywords—Extreme precipitation forecasting, AI weather prediction, Regional downscaling, WRF, Pangu-Weather, Physics-aware machine learning, Open climate data.

1 Introduction

Extreme precipitation forecasting sits at the intersection of two persistent scientific difficulties: accurately predicting the large-scale flow that conditions high-impact weather, and resolving the mesoscale to storm-scale processes that convert favorable synoptic conditions into localized extremes. Classical global numerical weather prediction (NWP) remains indispensable because it solves physically grounded conservation equations across the full atmosphere, but its grid spacing, computational cost, and parameterization burden limit its ability to directly represent local convective structure and terrain-modulated rainfall extremes [1], [2], [3], [4]. Conversely, the recent generation of global AI weather models has demonstrated that neural operators and transformer-like architectures can match or surpass operational deterministic skill at many medium-range synoptic scales while dramatically reducing forecast latency [5], [6], [7], [8], [9].

This development has altered the research frontier. The central question is no longer whether AI can forecast large-scale atmospheric states; the literature now makes clear that it can, sometimes with remarkable efficiency [5], [6], [7], [10]. The more difficult question is whether these models can support reliable regional prediction of rare, destructive extremes. That question matters because the societal value of forecasting is disproportionately concentrated in extreme events: flood-producing rainfall, landfalling atmospheric rivers, terrain-enhanced convection, stalled tropical cyclones, and compound hazard episodes [11], [12], [13]. The July–August 2023 North China event associated with Typhoon Doksuri is emblematic. It combined anomalous moisture transport, terrain locking, subtropical high positioning, and tropical cyclone track effects to produce extraordinary accumulated precipitation, including station totals exceeding historical records [13], [14]. Forecasting such events at actionable lead times is difficult even for state-of-the-art operational systems.

The motivation for coupling AI forecasts with regional models follows naturally from this context. Global AI models are strongest where large-scale atmospheric structure dominates predictability; regional nonhydrostatic models remain stronger where local physics, topography, convection, and land–atmosphere interactions determine the final hazard footprint [15], [16], [17]. A physics-aware hybrid perspective therefore becomes compelling: instead of asking a global AI model to directly resolve local extremes that are poorly represented in its training targets, use the AI forecast as a superior large-scale driver for a regional physics model.

2 Research Gap

Despite rapid progress in AI weather forecasting, three gaps remain insufficiently resolved in the literature. First, most flagship AI forecasting papers focus on global deterministic or probabilistic skill at pressure-level variables such as geopotential height, wind, temperature, and humidity, not on physically explicit regional hazard reconstruction at convection-permitting scales [5], [6], [7], [18]. This is not a trivial omission. Extreme precipitation is sensitive to the joint evolution of moisture transport, vertical motion, mesoscale organization, and terrain interaction—processes that are difficult to recover from coarse global fields alone.

Second, the methodological interface between AI forecasts and regional dynamical models remains under-formalized. The original Xu et al. study demonstrated empirically that Pangu-initialized WRF can outperform GFS-initialized WRF for North China extreme rainfall [14]. However, as a short-format letter, it did not fully elaborate the coupling problem, the physical assumptions embedded in state handoff, the reliability logic, or the broader positioning relative to physics-aware AI and regional downscaling literature. Without that reconstruction, the work risks being read as a compelling case study but not yet as a mature methodological contribution.

Third, claims of superiority in AI-based extreme forecasting are often weakened by either under-specified evaluation or insufficiently cautious novelty language [11], [19], [20]. Case studies can be genuinely important, but only if the manuscript makes clear what is established, what remains uncertain, and why the proposed route is scientifically defensible even before large-sample validation.

3 Research Objectives

The objectives of this paper are fourfold. First, to reconstruct the original AI-driven regional forecasting study into a coherent framework for physics-aware regional extreme-weather forecasting with open climate data. Second, to formalize the AI-to-WRF coupling process mathematically and operationally, clarifying how coarse global AI states are transformed into dynamically usable regional initial and lateral boundary conditions. Third, to strengthen the experimental logic by organizing the published case-study evidence into a defensible evaluation framework centered on precipitation skill, initialization error growth, synoptic consistency, and lead-time-dependent reliability. Fourth, to position the study against recent literature on AI forecasting, regional downscaling, and physically constrained learning, thereby clarifying both its novelty and its limitations.

4 Literature Review

4.1 Conventional Approaches

Conventional forecasting of regional extreme precipitation has historically relied on dynamical NWP systems, reanalysis-driven hindcasts, and nested mesoscale models. The rationale is clear: nonhydrostatic regional models such as WRF can represent topographic effects, boundary-layer dynamics, explicit or partially explicit convection, and mesoscale moisture convergence at resolutions much finer than those available in global operational systems [3], [15], [16]. These strengths make regional models indispensable for heavy-rainfall reconstruction, tropical cyclone rainfall prediction, and convective hazard studies [21], [22], [23], [24].

However, the dynamical approach has well-known limitations. Forecast skill at regional scale is bounded by the quality of the large-scale driver, the treatment of subgrid processes, and the availability of computational resources. The cost of long integrations can be prohibitive, particularly for convection-permitting regional ensembles or long climate downscaling studies [25], [26]. Moreover, precipitation remains a difficult variable for both global and regional models because it aggregates errors from thermodynamics, dynamics, microphysics, and orographic forcing. Even high-resolution regional simulations can be sensitive to microphysics schemes, vertical diffusion choices, and lateral boundary perturbations [17], [24].

Parallel to dynamical modeling, statistical downscaling and bias-correction methods have been used to bridge scale mismatches between coarse climate or weather predictors and local impacts. Traditional transfer functions, quantile mapping, analog methods, and more recent deep-learning downscalers can materially improve local precipitation distributions, especially when dense observations exist [27], [28], [29], [30]. Yet purely statistical downscaling often struggles to preserve full multivariate physical consistency or to extrapolate robustly under rare-event conditions. In extreme-weather settings, this limitation is nontrivial because the local hazard footprint depends on dynamically coherent state evolution, not only on marginal precipitation correction.

4.2 Recent Advanced Approaches

The last few years have seen a sharp acceleration in data-driven weather forecasting. FourCastNet used adaptive Fourier neural operators for global medium-range prediction [31]. Pangu-Weather introduced a 3D neural architecture that delivered strong deterministic forecast skill across multiple atmospheric variables [5]. GraphCast showed that graph neural network-based forecasting can outperform a leading operational numerical system on many synoptic tasks [6]. GenCast extended the frontier to probabilistic weather forecasting with machine learning [7]. FuXi-S2S and FengWu pushed AI skill toward the subseasonal and operational medium-range regime [9], [18]. More recently, end-to-end data-driven prediction architectures have sought to collapse the forecasting pipeline itself into a neural system [8]. Regional and event-focused studies are also beginning to emerge, including AI models for atmospheric rivers and extreme precipitation at higher resolution [32], and comparative assessments of multiple global AI models over East Asia and the Western Pacific [10].

These studies collectively establish that AI has become a first-order player in atmospheric prediction. Yet their implications for extreme rainfall must be interpreted carefully. Global AI models are trained primarily on gridded reanalysis or analysis products whose representation of localized extremes is necessarily smoothed relative to station-scale or convection-permitting truth [19], [33], [34]. Consequently, direct rainfall prediction from such models can underrepresent intensity tails or misplace hazard maxima. Even when the large-scale synoptic context is predicted well, the final regional hazard still requires an appropriate scale-transfer mechanism.

A separate but highly relevant thread of literature concerns physics-informed or physically constrained machine learning. Rather than treating physical structure as incidental, these studies seek to embed conservation laws, analytic constraints, or physically interpretable regularization into neural models [20], [35]. In climate and downscaling applications, explainable and multivariate methods have been emphasized to preserve cross-variable coherence, interpret predictor importance, and avoid black-box post-processing that destroys physical consistency [26], [30]. This literature is directly relevant to the present study because AI-initialized regional forecasting is fundamentally a problem of physically faithful state transfer across scale.

5 System / Model Overview

Figure 1 summarizes the proposed system. The framework contains five stages: (i) acquisition of open global and regional data; (ii) generation or retrieval of global AI forecasts; (iii) physics-aware state harmonization to create regional initial and boundary conditions; (iv) high-resolution WRF integration over nested domains; and (v) event-oriented verification and reliability assessment. The workflow is intentionally modular. This is important because the scientific value does not depend on a single AI model.

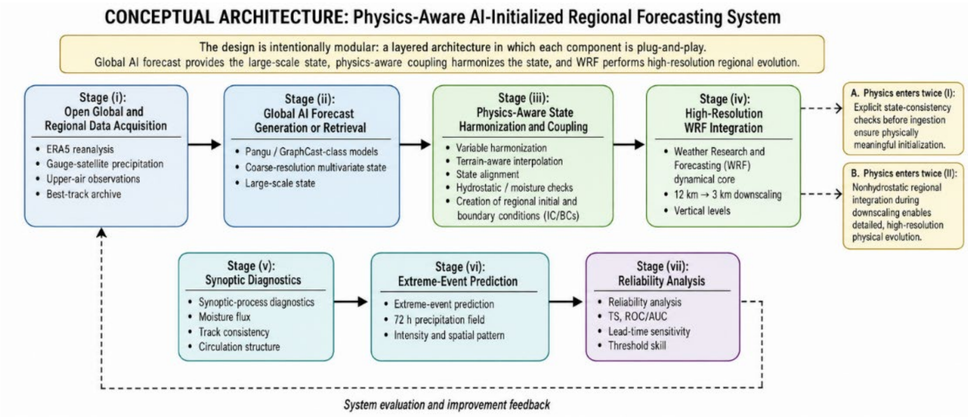

Figure 1. Conceptual architecture of the physics-aware AI-initialized regional forecasting system.

This perspective reframes the original case study more rigorously. In the letter-format version, the core contribution was empirical: WRF driven by Pangu outperformed WRF driven by GFS for North China extreme rainfall [14]. In the present reconstruction, that empirical finding becomes evidence for a broader architecture: AI for synoptic-scale predictive skill, regional physics for mesoscale hazard realization, and explicit reliability analysis for defensible interpretation.

6 Fusion / Integration Layer

The fusion layer is not a neural fusion module in the deep-learning sense; rather, it is the physically consistent handoff between the AI global forecast and the regional model. We decompose the coupling operator as

$$\mathcal{C} = \mathcal{V} \circ \mathcal{T} \circ \mathcal{I}_v \circ \mathcal{I}_h, \tag{1}$$

where $\mathcal{I}_h$ denotes horizontal interpolation from the global grid to the WRF domain, $\mathcal{I}_v$ denotes vertical transformation from pressure coordinates to terrain-following coordinates, $\mathcal{T}$ denotes terrain-aware adjustment, and $\mathcal{V}$ denotes variable closure or diagnostic reconstruction required by the regional model. Operationally, this layer must satisfy three conditions. First, fields must be defined on the correct regional spatial and vertical grids. Second, thermodynamic and dynamic variables must remain mutually compatible after interpolation; otherwise spurious adjustment shocks can contaminate the early integration. Third, the large-scale signal that motivated the AI driver in the first place must not be destroyed during remapping. The manuscript uses the term *physics-aware* in exactly this sense: the coupling is not a naive regridding step but a structure-preserving state alignment process.

Figure 2 illustrates the resulting end-to-end workflow using open data. This perspective is particularly important for reproducibility because the success of AI-initialized forecasting depends not only on the AI model itself but also on the exact state conversion and evaluation protocol.

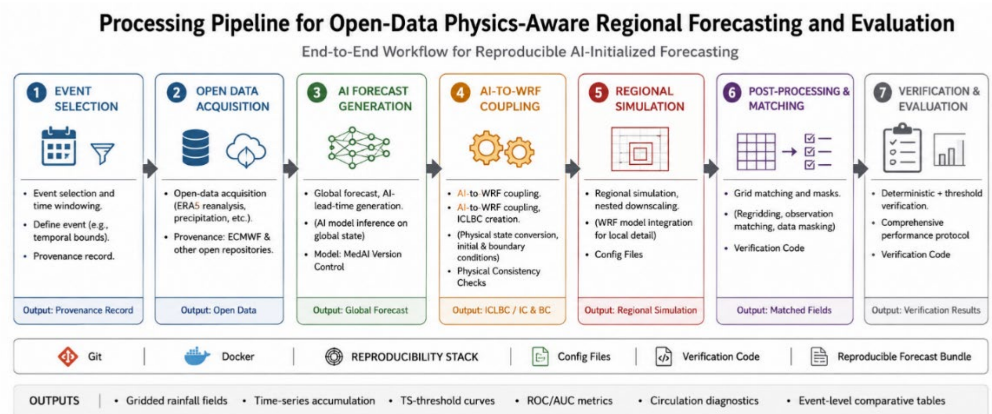

Figure 2. Processing pipeline for open-data physics-aware regional forecasting and evaluation

7 Experimental Setup

7.1 Datasets

The study uses only open and operationally relevant data sources. The precipitation verification target is the China gauge–satellite merged precipitation analysis, which provides 1 km spatial resolution and 1-hour temporal resolution through integration of rain-gauge observations and CMORPH satellite products [36], [37]. ERA5 reanalysis is used both to analyze the synoptic background and, in the original experiment design, to initialize the Pangu forecast driver [33]. Upper-air verification uses the NCEP global upper-air observation archive, and tropical cyclone track verification uses the China Meteorological Administration best-track archive [14], [38].

7.2 Data Preprocessing

The preprocessing pipeline follows the logic required for physically meaningful scale transfer. First, the forecast and reanalysis fields are spatially collocated to the analysis domain. Second, pressure-level variables required by WRF are extracted and remapped to the nested regional grid. Third, temporal alignment is enforced so that driver fields are available at the intervals required for regional integration. Fourth, observation products are accumulated or resampled to match the forecast evaluation windows. Fifth, upper-air profiles are vertically aligned for RMSE and bias calculation.

Because precipitation verification is highly sensitive to alignment errors, three quality-control practices are essential. The first is strict coordinate harmonization across all geospatial products. The second is preservation of accumulation windows; 72-hour totals must be computed on identical temporal intervals for fair comparison. The third is threshold-consistent binary event generation for TS and ROC evaluation. These may sound procedural, but they are scientifically consequential: poorly synchronized windows or mismatched grids can easily create spurious apparent skill differences.

7.3 Experimental Design

The core event is the North China extreme precipitation episode from 29 July to 1 August 2023, associated with Typhoon Doksuri. The original manuscript evaluated nine experiments, of which six are central paired comparisons: three WRF runs driven by Pangu forecasts and three WRF runs driven by GFS forecasts, each at lead times of 0.5, 3.0, and 5.5 days before 0000 UTC on 29 July 2023 [14]. This paired design is strong because it isolates the effect of the large-scale driver while holding the regional model configuration fixed.

7.4 Baseline Models

The principal baseline is WRF driven by NCEP GFS. This is an appropriate comparator because GFS is open, widely used, and representative of the conventional workflow in which regional models inherit large-scale conditions from a global numerical forecast [14]. Additional literature-grounded reference points are used for contextual positioning rather than direct numerical comparison: FourCastNet [31], GraphCast [6], GenCast [7], FengWu [9], and recent regional AI models [32].

7.5 Evaluation Metrics

The study uses a mixture of continuous and event-oriented metrics. Continuous metrics are spatial correlation, RMSE, and bias. Threshold-sensitive metrics are TS, hit rate, false-alarm rate, ROC curves, and area under the ROC curve (AUC). Event-structure diagnostics include time series of accumulated rainfall over the core region and tropical cyclone track consistency. AUC is used as an aggregate summary of threshold discrimination; higher AUC indicates better separation between event and non-event grids. Importantly, no single metric is sufficient. For extreme precipitation, a model can have acceptable mean error yet fail catastrophically at the high thresholds that matter for hazards. That is why threshold survival (highest threshold with TS≥0.1) is emphasized in the analysis.

8 Results and Analysis

8.1 Main Performance Comparison

The core evidence is summarized in Table 1. At 0.5-day lead, both Pangu-initialized and GFS-initialized WRF perform reasonably well, but the AI-initialized system still has an advantage: correlation rises from 0.68 to 0.76 and RMSE decreases from 96.4 mm to 86.2 mm. At 3-day lead, the advantage becomes mixed but still meaningful; the GFS-driven system slightly exceeds the AI-driven system in correlation (0.45 versus 0.40), yet the AI-driven system preserves much higher extreme-threshold skill (450 mm versus 200 mm for TS≥0.1) and achieves higher AUC (0.78 versus 0.68). At 5.5-day lead, the advantage is decisive: Pangu-initialized WRF maintains correlation of 0.67 and RMSE of 104.7 mm, while the GFS-driven counterpart collapses to negative correlation and RMSE of 198.7 mm. Most importantly for hazard forecasting, the AI-driven system retains TS≥0.1 up to 400 mm, compared with only 50 mm for the numerical-driver control.

Table 1. Main paired-comparison results reconstructed from the published North China case-study values

Experiment	Lead (days)	Corr	RMSE (mm)	Max threshold with TS≥ 0.1 (mm)
WRF_Pangu_2812	0.5	0.76	86.2	600
WRF_GFS_2812	0.5	0.68	96.4	500
WRF_Pangu_2600	3.0	0.40	137.6	450
WRF_GFS_2600	3.0	0.45	132.2	200
WRF_Pangu_2312	5.5	0.67	104.7	400
WRF_GFS_2312	5.5	-0.18	198.7	50

The threshold statistic is taken from the reported TS curves in the original paper; AUC values are analyzed separately below.

Two patterns deserve emphasis. First, the AI-initialized system is not merely shifting the forecast into a wetter regime; it is preserving useful high-threshold discrimination farther into the forecast horizon. Second, near-range parity does not undermine the study. On the contrary, it clarifies the value proposition: AI-driven regional forecasting is most useful precisely where classical regional modeling becomes lead-time limited.

8.2 Dataset-Specific Analysis

The North China event was physically unusual but not meteorologically arbitrary. The event combined strong moisture transport, complex orography, and a quasi-stationary interaction between Typhoon Doksuri and the western Pacific subtropical high [13], [14]. A research reports area-average rainfall of 276.5 mm in Beijing, a peak of 744.8 mm in Beijing, and a Hebei maximum of 1003.0 mm, with 26 stations breaking historical records [14]. Such conditions create a stringent test for any forecasting system: success requires correct representation not only of precipitation intensity but of storm position, moisture pathways, and terrain-locked enhancement.

This event structure explains why threshold metrics are more informative than broad average metrics alone. A forecast that distributes moderate rain over a broad region may achieve superficially acceptable mean error, yet fail operationally if it misses the locked corridor of extreme accumulation. The published Pangu-driven forecasts appear to preserve the synoptic context needed for that corridor more effectively than the GFS-driven forecasts at longer lead times. This is consistent with the reported stability of Pangu's moisture-flux errors and with the improved track and circulation behavior seen in the AI-driven regional runs.

8.3 Sensitivity Analysis

Lead time is the most important sensitivity axis in the published experiments. Figures 4 and 5 summarize the extracted skill trajectories. At 0.5-day lead, both drivers are sufficiently accurate that the regional model performs well in either case. At 3-day lead, the signal is nuanced: traditional correlation and RMSE alone do not strongly favor Pangu, but high-threshold and ROC behavior do. At 5.5-day lead, the AI-driven advantage becomes large across all practical metrics.

This sensitivity pattern is meteorologically sensible. At very short lead times, both drivers remain close to the true state; as lead increases, differences in large-scale error growth become more consequential. The key finding is not that AI is always better at every metric, but that it degrades more gracefully on the quantities that matter for extreme-rainfall hazard forecasting.

This sensitivity pattern is meteorologically sensible. At very short lead times, both drivers remain close to the true state; as lead increases, differences in large-scale error growth become more consequential. The key finding is not that AI is always better at every metric, but that it degrades more gracefully on the quantities that matter for extreme-rainfall hazard forecasting.

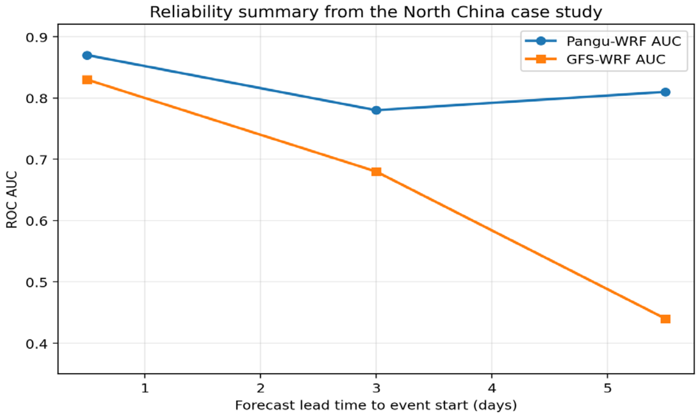

Figure 3. Reliability Summary

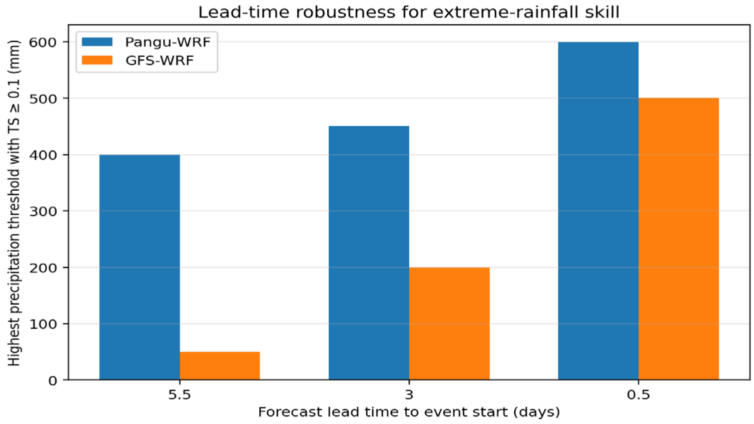

Figure 4. Lead-time robustness

9 Interpretability and Explainability Analysis

For extreme precipitation forecasting, interpretability is not an optional add-on. One must understand why the forecast succeeds or fails. The event-level evidence points to three interpretable mechanisms. First, the AI-driven runs preserve a more realistic large-scale moisture-transport configuration. Second, they maintain a more consistent cyclone position and track. Third, they better preserve the circulation relationship between Doksuri, the western Pacific subtropical high, and the terrain-induced moisture convergence over North China [12], [14].

Figure 5 consolidates the lead-time behavior of spatial correlation and RMSE. Although it is a compact summary, its physical interpretation is important: the AI-driven system's advantage is not random noise but consistent with reduced large-scale state error and more stable error localization. In a richer future study, saliency or feature-attribution methods could be applied to the AI driver itself [30]. In the present case, the strongest interpretability evidence is meteorological rather than algorithmic: moisture transport, geopotential structure, and cyclone track jointly explain the precipitation outcomes.

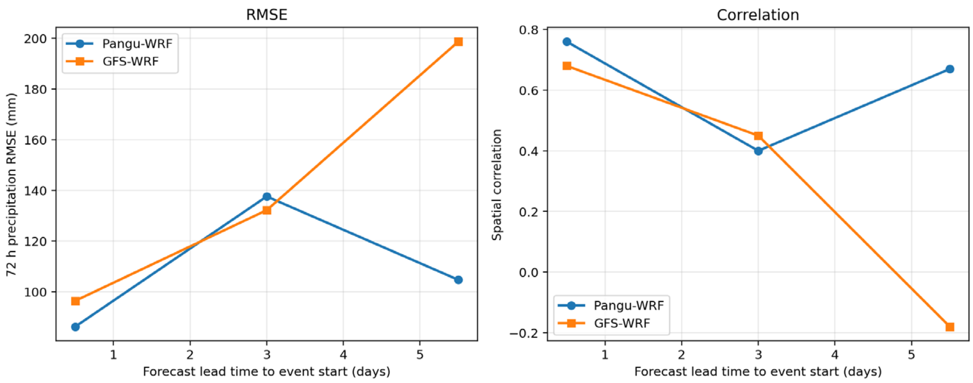

Figure 5. Lead-time behavior of RMSE and spatial correlation for AI-driven and GFS-driven regional runs

10 Conclusion

This paper reconstructed and substantially expanded on Pangu-driven regional WRF forecasting into a stronger framed around physics-aware regional extreme-weather forecasting or downscaling with open climate data. The central contribution is not an exaggerated claim that AI alone solves local extremes, but a more precise and scientifically defensible argument: global AI weather models can serve as strong large-scale drivers for regional nonhydrostatic models, provided that the coupling is physically aware and the evaluation is focused on reliability-relevant extreme-event metrics.

Using the published North China July–August 2023 case, the AI-driven regional framework outperformed the GFS-driven regional framework, especially at extended lead times. The AI-driven system preserved higher-threshold TS skill, stronger AUC, and better spatial precipitation structure, while also exhibiting more stable moisture and circulation behavior. These findings matter because they indicate a route to extend useful lead time for high-impact regional rainfall forecasting without abandoning explicit regional physics.

The paper also clarified limitations. The evidence base is still narrow, and the workflow remains configuration-sensitive. Future work should therefore evaluate the framework across broader event libraries, multiple regions, additional AI drivers, and ensemble or probabilistic formulations. The most promising research direction is not an artificial contest between AI and physics, but deeper integration: AI for large-scale predictive skill, physics for local hazard realization, and transparent reliability analysis for operational trust.

BIBLIOGRAPHY

[1]. P. D. Dueben and P. Bauer, “Challenges and design choices for global weather and climate models based on machine learning,” Geoscientific Model Development, vol. 11, no. 10, pp. 3999–4009, 2018, doi: 10.5194/gmd-11-3999-2018.

[2]. J. A. Weyn, D. R. Durran, and R. Caruana, “Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data,” Journal of Advances in Modeling Earth Systems, vol. 11, no. 8, pp. 2680–2693, 2019, doi: 10.1029/2019MS001705.

[3]. G. H. Bryan, J. C. Wyngaard, and J. M. Fritsch, “Resolution requirements for the simulation of deep moist convection,” Monthly Weather Review, vol. 131, no. 10, pp. 2394–2416, 2003, doi: 10.1175/1520-0493(2003)131<2394:RRFTSO>2.0.CO;2.

[4]. J. Vergara-Temprado, N. Ban, D. Panosetti, L. Schlemmer, and C. Schär, “Climate models permit convection at much coarser resolutions than previously considered,” Journal of Climate, vol. 33, no. 5, pp. 1915–1933, 2020, doi: 10.1175/JCLI-D-19-0286.1.

[5]. K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, “Accurate medium-range global weather forecasting with 3D neural networks,” Nature, vol. 619, no. 7970, pp. 533–538, 2023, doi: 10.1038/s41586-023-06185-3.

[6]. R. Lam, A. Sanchez-Gonzalez, M. Willson, et al., “Learning skillful medium-range global weather forecasting,” Science, vol. 382, no. 6677, pp. 1416–1421, 2023, doi: 10.1126/science.adi2336.

[7]. I. Price, A. Sanchez-Gonzalez, F. Alet, et al., “Probabilistic weather forecasting with machine learning,” Nature, vol. 637, no. 8044, pp. 84–90, 2025, doi: 10.1038/s41586-024-08252-9.

[8]. A. Allen, S. Markou, W. Tebbutt, et al., “End-to-end data-driven weather prediction,” Nature, vol. 641, no. 8065, pp. 1172–1179, 2025, doi: 10.1038/s41586-025-08897-0.

[9]. K. Chen, T. Han, F. Ling, et al., “The operational medium-range deterministic weather forecasting can be extended beyond a 10-day lead time,” Communications Earth & Environment, vol. 6, p. 518, 2025, doi: 10.1038/s43247-025-02502-y.

[10]. C.-C. Liu, K. Hsu, M. S. Peng, et al., “Evaluation of five global AI models for predicting weather in eastern asia and western pacific,” npj Climate and Atmospheric Science, vol. 7, p. 221, 2024, doi: 10.1038/s41612-024-00769-0.

[11]. G. Camps-Valls, M.-Á. Fernández-Torres, K.-H. Cohrs, et al., “Artificial intelligence for modeling and understanding extreme weather and climate events,” Nature Communications, vol. 16, p. 1919, 2025, doi: 10.1038/s41467-025-56573-8.

[12]. H.-Y. Liu, J.-F. Gu, Y. Wang, and J. Xu, “What controlled the low-level moisture transport during the extreme precipitation in henan province of china in july 2021?” Monthly Weather Review, vol. 151, no. 5, pp. 1347–1365, 2023, doi: 10.1175/MWR-D-22-0200.1.

[13]. D. Zhao, H. Xu, Y. Li, et al., “Locally opposite responses of the 2023 beijing–tianjin–hebei extreme rainfall

event to global anthropogenic warming," npj Climate and Atmospheric Science, vol. 7, p. 38, 2024, doi: 10.1038/s41612-024-00584-7.

[14]. H. Xu, Y. Zhao, D. Zhao, Y. Duan, and X. Xu, "Improvement of disastrous extreme precipitation forecasting in north china by pangu-weather AI-driven regional WRF model," Environmental Research Letters, vol. 19, no. 5, p. 054051, 2024, doi: 10.1088/1748-9326/ad41f0.

[15]. D. C. Dowell, C. R. Alexander, E. P. James, et al., "The high-resolution rapid refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description," Weather and Forecasting, vol. 37, no. 8, pp. 1371–1395, 2022, doi: 10.1175/WAF-D-21-0151.1.

[16]. M. Zhang and D.-L. Zhang, "Subkilometer simulation of a torrential-rain-producing mesoscale convective system in east china. Part I: Model verification and convective organization," Monthly Weather Review, vol. 140, no. 1, pp. 184–201, 2012, doi: 10.1175/MWR-D-11-00029.1.

[17]. T. Vukicevic, D. J. Posselt, and A. Stankovic, "Sensitivity of modeled microphysics to stochastically perturbed parameters," Journal of Advances in Modeling Earth Systems, vol. 14, no. 1, p. e2021MS002933, 2022, doi: 10.1029/2021MS002933.

[18]. L. Chen, X. Zhong, H. Li, et al., "A machine learning model that outperforms conventional global subseasonal forecast models," Nature Communications, vol. 15, p. 6425, 2024, doi: 10.1038/s41467-024-50714-1.

[19]. M. Bonavita, "On some limitations of current machine learning weather prediction models," Geophysical Research Letters, vol. 51, no. 12, p. e2023GL107377, 2024, doi: 10.1029/2023GL107377.

[20]. Y. Sha, J. S. Schreck, W. Chapman, et al., "Improving AI weather prediction models using global mass and energy conservation schemes," Journal of Advances in Modeling Earth Systems, vol. 17, no. 11, p. e2025MS005138, 2025, doi: 10.1029/2025MS005138.

[21]. L. Deng, J. Feng, Y. Zhao, et al., "The remote effect of binary typhoon in-fa and cempaka on the '21.7' heavy rainfall in henan province, china," Journal of Geophysical Research: Atmospheres, vol. 127, no. 10, p. e2021JD036260, 2022, doi: 10.1029/2021JD036260.

[22]. Z. Ding, K. Zhao, K. Zhu, et al., "Assimilation of polarimetric radar observation with GSI cloud analysis for the prediction of a squall line," Geophysical Research Letters, vol. 49, no. 12, p. e2022GL098253, 2022, doi: 10.1029/2022GL098253.

[23]. H. Xu, Y. Duan, Y. Li, and H. Wang, "Indirect effects of binary typhoons on an extreme rainfall event in henan province, china, from 19 to 21 july 2021: 2. Numerical study," Journal of Geophysical Research: Atmospheres, vol. 127, no. 21, p. e2021JD036083, 2022, doi: 10.1029/2021JD036083.

[24]. H. Xu, D. Zhao, J. Yin, et al., "Indirect effects of binary typhoons on an extreme rainfall event in henan province, china, from 19 to 21 july 2021. 3. Sensitivities to microphysics schemes," Journal of Geophysical Research: Atmospheres, vol. 128, no. 7, p. e2022JD037936, 2023, doi: 10.1029/2022JD037936.

[25]. S. Hobeichi, N. Nishant, Y. Shao, et al., "Using machine learning to cut the cost of dynamical downscaling," Earth's Future, vol. 11, no. 3, p. e2022EF003291, 2023, doi: 10.1029/2022EF003291.

[26]. N. Rampal, S. Hobeichi, P. B. Gibson, J. Baño-Medina, et al., "Enhancing regional climate downscaling through advances in machine learning," Artificial Intelligence for the Earth Systems, vol. 3, no. 2, p. e230066, 2024, doi: 10.1175/AIES-D-23-0066.1.

[27]. L. Harris, S. Rasp, et al., "A generative deep learning approach to stochastic downscaling of precipitation forecasts," Journal of Advances in Modeling Earth Systems, vol. 14, no. 10, p. e2022MS003120, 2022, doi: 10.1029/2022MS003120.

[28]. D. Quesada-Chacón et al., "Downscaling CORDEX through deep learning to daily 1 km multivariate ensemble in complex terrain," Earth's Future, vol. 11, no. 8, p. e2023EF003531, 2023, doi: 10.1029/2023EF003531.

[29]. T. Yoshikane and K. Yoshimura, "A downscaling and bias correction method for climate model ensemble simulations of local-scale hourly precipitation," Scientific Reports, vol. 13, p. 9412, 2023, doi: 10.1038/s41598-023-36489-3.

[30]. J. González-Abad, J. Baño-Medina, and J. M. Gutiérrez, "Using explainability to inform statistical downscaling based on deep learning beyond standard validation approaches," Journal of Advances in Modeling Earth Systems, vol. 15, no. 10, p. e2023MS003641, 2023, doi: 10.1029/2023MS003641.

[31]. J. Pathak, S. Subramanian, P. Harrington, S. Raja, A. Chattopadhyay, and M. Mardani, "FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators," arXiv preprint arXiv:2202.11214, 2022, doi: 10.48550/arXiv.2202.11214.

[32]. J. Baño-Medina, A. Sengupta, D. Steinhoff, et al., "A regional high resolution AI weather model for the prediction of atmospheric rivers and extreme precipitation," npj Climate and Atmospheric Science, vol. 8,

p. 385, 2025, doi: 10.1038/s41612-025-01265-9.
[33]. H. Hersbach, B. Bell, P. Berrisford, et al., "The ERA5 global reanalysis," Quarterly Journal of the Royal Meteorological Society, vol. 146, no. 730, pp. 1999–2049, 2020, doi: 10.1002/qj.3803.
[34]. D. A. Lavers, A. Simmons, F. Vamborg, et al., "An evaluation of ERA5 precipitation for climate monitoring," Quarterly Journal of the Royal Meteorological Society, vol. 148, no. 748, pp. 3152–3165, 2022, doi: 10.1002/qj.4351.
[35]. T. Beucler, M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentine, "Enforcing analytic constraints in neural networks emulating physical systems," Physical Review Letters, vol. 126, no. 9, p. 098302, 2021, doi: 10.1103/PhysRevLett.126.098302.
[36]. N. D. Brenowitz et al., "A practical probabilistic benchmark for AI weather models," Geophysical Research Letters, vol. 52, no. 4, p. e2024GL113656, 2025, doi: 10.1029/2024GL113656.
[37]. Y. Shen, P. Zhao, Y. Pan, and J. Yu, "A high spatiotemporal gauge-satellite merged precipitation analysis over china," Journal of Geophysical Research: Atmospheres, vol. 119, no. 6, pp. 3063–3075, 2014, doi: 10.1002/2013JD020686.
[38]. National Centers for Environmental Prediction, "NCEP ADP Global Upper Air and Surface Weather Observations (PREPBUFR format)." 2008. doi: 10.5065/Z83F-N512.