https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

SPIN (v1.0): A Spontaneous Synthetic Tropical Cyclone Model Empowered by NeuralGCM for Hazard Assessment

Yurong Gao[1], Dazhi Xi[1]

[1]Department of Earth and Planetary Sciences, The University of Hong Kong, Hong Kong, 0000, China

Correspondence: Dazhi Xi (dazhixi@hku.hk)

Abstract. A hybrid framework for simulating SPontaneous synthetic tropical cyclones (TCs) with realistic INtensity, hereafter SPIN, is developed for TC risk assessment. A key advantage of SPIN over previous synthetic TC models is that it avoids the assumption of independence between TCs, while enabling two-way interactions between synthetic TCs and their ambient environment. The SPIN model leverages a Neural General Circulation Model (NeuralGCM) to simulate spontaneously generated TC tracks, and then couples a dynamic TC intensity model to estimate their intensity evolutions based on the large-scale environment. SPIN reproduces the observed climatology of TC activity, including interannual variability, seasonal cycle, genesis, tracks, and lifetime maximum intensity distributions. It also faithfully reproduces the observed return periods of landfall intensity across different regions, enabling its future application to TC risk assessment. Beyond individual TC events, SPIN demonstrates improved skills in representing multiple tropical cyclone events (MTCEs), including their interannual variability, peak concurrent TC count per cluster, and the spatial relationship between consecutive TCs. By circumventing the independent TC assumption and allowing for two-way TC-environment interactions, SPIN opens new potential for assessing compound hazards like MTCE and beyond.

1 Introduction

Tropical cyclones (TCs) are among the most destructive hazards in coastal regions. On average, about 90 TCs form worldwide each year (Emanuel, 2006; Zhou and Lin, 2024), affecting 20.4 million people annually and causing mean direct economic losses of 51.5 billion USD over the past decade (Krichene et al., 2023). Beyond the impacts of individual storms, back-to-back TCs (Fu et al., 2025; Xi et al., 2023; Xi and Lin, 2021) can lead to especially severe impact, as communities may not recover from one event before the next occurs (Zscheischler et al., 2018). These societal consequences underscore the importance of assessing not only the risks of individual TCs but also those arising from their clustering.

The high risks associated with TCs lead to the development of TC hazard downscaling frameworks. Over recent decades, two major approaches have been used to downscale TC hazards: statistical and dynamical downscaling. Statistical downscaling models (Bloemendaal et al., 2020; Emanuel et al., 2006; Jing and Lin, 2020; Lee et al., 2018; Lin et al., 2023) provide an efficient way to simulate TCs under given climate conditions. These models are driven by environmental variables and can rapidly generate large numbers of synthetic TCs to estimate the occurrence chances of the extremes, costing minimal

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

computational resources. A common drawback of these models is that the generated TCs do not feedback to the ambient environment. This limitation prevents the representation of two-way interactions between storms and their surroundings. Furthermore, all TCs in these models are assumed to be independent, which is not physically realistic, since storm–storm interactions do occur in nature. One prominent example is multiple TC events (MTCEs), where two or more TCs occur simultaneously within the same basin (Fu et al., 2023; Schenkel, 2016), with subsequent TCs potentially being preconditioned by synoptic-scale cyclonic disturbances generated by preceding TCs (Ritchie and Holland, 1999; Yoshida and Ishikawa, 2013). In contrast, dynamical downscaling explicitly simulates TCs in numerical models, allowing two-way interactions between the TCs and their environment. However, even the high-resolution models with grid spacing of 4–25 km still have difficulty in representing the most intense TCs (Buonomo et al., 2024; Davis, 2018; Judt et al., 2021). Most importantly, they are prohibitively expensive to run over long periods to produce the large samples needed for robust TC risk assessment.

The recent development of artificial intelligence (AI) weather forecast or climate models (Bi et al., 2023; Bodnar et al., 2025; Kochkov et al., 2024; Price et al., 2025) have made the rapid simulation of synthetic TCs that can interact with their surrounding environment possible. Building on this progress, Jing et al. (2024) developed a TC downscaling model based on a data-driven approach that accounts for two-way storm–environment interactions throughout the storm life cycle. Nonetheless, this approach still requires manually seeded vortices rather than permitting spontaneous genesis from environmental fields. In addition, as with numerical models, TCs directly extracted from AI-simulated outputs still tend to underestimate intensity, since the models are trained on ERA5 reanalysis data whose resolution cannot adequately resolve TC structure (Jing et al., 2024). In short, there remains a gap for models that can efficiently generate large storm samples spontaneously emerging from the environment, co-evolving with it, and achieving realistic maximum intensities.

Rather than functioning like most AI weather forecast models, the Neural General Circulation Model (NeuralGCM) (Kochkov et al., 2024) offers sufficient stability for long-term climate simulations by preserving a GCM-like numerical solver for large-scale dynamics while replacing small-scale parameterization schemes with a neural network. It reproduces key climate diagnostics with considerable skill, including realistic TC frequency and trajectories. Recent studies have further demonstrated its ability in predicting seasonal TC activity (Zhang et al., 2025), long-term heatwave projections (Duan et al., 2025), and the spectra of large-scale tropical waves (Baxter et al., 2025). These features suggest that NeuralGCM provides a suitable basis for developing a synthetic TC downscaling framework.

In this study, we propose a hybrid framework that functions as a synthetic TC downscaling model, termed SPIN (a SPontaneous synthetic tropical cyclone model with INtensity fidelity). The model employs a two-step process: firstly, it leverages NeuralGCM (Kochkov et al., 2024) to simulate ensembles of spontaneously generated TC tracks; then, it couples these tracks with the FAST TC intensity simulator (Emanuel, 2017; Lin et al., 2023) to compute realistic intensities based on the large-scale environment. Compared with traditional statistical synthetic TC models, this model provides a more physically consistent representation of synthetic TCs and their ambient environment by generating TCs that emerge from and interact

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

with the environment, while remaining computationally more efficient than dynamical downscaling. Thus, this model has proved to be particularly advantageous for the study of compound TC hazards such as MTCEs.

This paper is organized as follows. Section 2 describes the data used for model development along with the model components and evaluation methods. Section 3 evaluates SPIN simulated climatological characteristics of TCs and MTCEs against observations. Section 4 compares observed and simulated return periods at different regions of the globe. Section 5 provides a summary and discussion.

2 Materials and Methods

2.1 Observational and Reanalysis Data

The ERA5 reanalysis data (Hersbach et al., 2020) from 1979 to 2022 at 0.25° × 0.25° resolution are used to initialize and force the NeuralGCM component of SPIN (see Section 2.2). The initialization fields include wind, geopotential, temperature, specific humidity, and specific cloud ice and liquid water content on pressure levels, while the forcing consists of sea surface temperature (SST) and sea ice concentration (SIC) at the surface level. The simulated TCs are evaluated at both global and basin scales. We adopt a similar basin classification as Lin et al. (2023), dividing the global ocean into seven basins where TCs are prevalent: the Western Pacific (WP), Eastern Pacific (EP), Northern Atlantic (NA), Northern Indian Ocean (NI), Southern Indian Ocean (SI), Southern Pacific (SP), and the Australian Basin (AU). The boundaries of each basin are illustrated in Figure S1. For model evaluation, we use the USA archive from the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010), which provides 3-hourly latitude and longitude positions as well as maximum sustained wind speeds for each storm.

2.2 Model Components

The SPIN model follows a two-step approach. First, it leverages NeuralGCM (Kochkov et al., 2024) to simulate ensembles of spontaneously generated TC tracks. Then, FAST intensity model (Emanuel, 2017; Lin et al., 2023) is coupled to each track to compute realistic intensities based on the large-scale environment. In this way, SPIN generates large samples of TCs that emerge from and interact with the simulated environment.

2.2.1. Simulation of Environmental Fields

NeuralGCM is a hybrid physics-machine learning atmospheric model (Kochkov et al., 2024) that preserves a GCM-like numerical solver for large-scale dynamics, while replacing small-scale parameterization schemes with a neural network. It offers orders-of-magnitude computational savings over conventional GCMs and is sufficiently stable for long-term climate simulations. It has been shown to reproduce key climate diagnostics with considerable skill, and demonstrates promising performance in simulating realistic TC activity in tests for the year 2020 (Kochkov et al., 2024).

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

We use the pre-trained 1.4° deterministic version of NeuralGCM to simulate the environmental fields from which TC tracks are extracted. The 1.4° version is about ten times faster in inference than the 0.7° version while still producing reliable TC activity, and was also adopted in Kochkov et al. (2024). Following Kochkov et al. (2024), we initialize a 14-member ensemble at 6-hour intervals in mid-October of each year from 1979 to 2021. Each ensemble member is integrated from its initialization date through December of the following year, with outputs saved every 6 hours. Simulations are then evaluated for the January-December periods from 1980 to 2022. Monthly SST and SIC fields are linearly interpolated to 12-hourly resolution and updated accordingly to force the model throughout the simulation.

2.2.2. Tracking of TCs

We apply the TempestExtreme method (Ullrich et al., 2021) to track TCs from the simulated environmental fields, using geopotential and vorticity as primary criteria, and without imposing any minimum wind speed thresholds. In this study, most parameter settings in TempestExtreme are similar with those in Kochkov et al. (2024), who ensures that the global frequency of identified TCs matches the counts obtained from the 0.25° × 0.25° ERA5 reanalysis data. We further fine-tune parameters separately for each basin, aiming to align the simulated historical annual average number of tracks and storm lifetimes with those reported in the IBTrACS. A total of 14 ensemble members, each simulated for 43 years, produces approximately 52,000 tracks globally. The ensemble mean number of extracted TCs in each basin agrees well with the observations, without any additional bias correction (Figure S2).

2.2.3. Intensity model

TC intensities along the modeled tracks are simulated with the FAST intensity model, originally proposed by Emanuel (2017) and Emanuel and Zhang (2017), and later extended to the global scale by Lin et al. (2023). FAST is a physically based and computationally efficient model that simulates the maximum sustained wind speed by solving two coupled ordinary differential equations. This model has been successfully used in probabilistic TC forecasting (Lin et al., 2020) and in physics-based statistical downscaling frameworks (Lin et al., 2023).

In the Lin et al. (2023; hereafter JL23) framework, only tracks that reach the intensity threshold of 18 m s^{-1} within 48 hours are retained, while others are discarded. In our approach, we retain all extracted tracks to maintain consistency with the observed track count. For tracks that do not reach 18 m s^{-1} within 48 hours, we apply a bias correction at the 48-hour point and all subsequent times by adding a random constant offset, so that the intensity of 48-hour is at or slightly above 18 m s^{-1}, following the protocol of Lin et al. (2023). All model parameters are adopted from Lin et al. (2023), except for the surface enthalpy coefficient (Ck), which is regionally tuned for the assessment of lifetime maximum intensity (LMI) and landfall intensity. Because the FAST equations require initialization of the maximum azimuthal wind speed using a random seeding approach, we perform 10 simulations per track with different random seeds, resulting in 10 intensity realizations for each

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

ensemble member (equivalent to 6020-year of simulation in total). In our analysis, we define TC genesis as the first time the storm reaches 18 m s^{-1}, and dissipation as the last time it exceeds this threshold.

2.3. Model Evaluation and Benchmarks

2.3.1. Benchmark models

To assess the intensity component of SPIN, we benchmark against the physics-based statistical downscaling framework developed by JL23, which applies the FAST model to statistical synthetic tracks.

For comparison of SPIN-simulated MTCE (see Section 2.5 for definition) metrics with statistical downscaling methods, we use two benchmarks: (1) the JL23 statistical downscaling framework, and (2) a monthly-uniform-sampled version of IBTrACS. Like most statistical downscaling models, the JL23 model reports tracks on a monthly scale rather than giving precise genesis and dissipation dates. We therefore assign dates by uniformly sampling within each month for the tracks produced by the JL23 model, following previous studies (Xi et al., 2023), and perform 1000 Monte Carlo simulations of the daily-scale genesis dates to quantify uncertainty. The number of sampled tracks per basin in JL23 follows its own TC frequency. This uniform-sampling strategy assumes independent TC genesis and neglects storm-storm interactions. To circumvent the bias in TC genesis frequency in JL23, we also apply the same sampling approach to observed TCs from IBTrACS dataset. Comparing SPIN with these benchmarks, we test whether SPIN can more faithfully reproduce observed MTCE characteristics than a month-level uniform-sampling framework (both benchmarks).

2.3.2. Statistical Metrics

To quantify SPIN performance compared to the benchmarks in simulating observed probability distributions of various TC-related metrics (e.g. LMI, TC inter-genesis dates), we use two statistical measures: the Kullback-Leibler (KL) divergence and the area between the cumulative distribution functions (ACDFs) of the model and observations.

The KL divergence $D_{\mathrm{KL}}(M \parallel O)$ measures the difference between the probability distributions of model (M) and observations (O) for a given metric, reflecting the information loss when the model distribution is used to approximate the observations. It is particularly sensitive to regions where the model assigns low probability to events that are frequent in observations, and is therefore especially useful for evaluating differences in the tails of distributions. It is defined as:

$$D_{KL}(M \parallel O) = \sum_i P_M(i) \log\left(\frac{P_M(i)}{P_O(i)}\right) \quad (1)$$

where $P_M(i)$ and $P_O(i)$ denote the probability of the i-th bin for the model and observations, respectively. The value of $D_{\mathrm{KL}}(M \parallel O) = 0$ indicates perfect agreement between the two distributions. Higher values indicate greater divergence.

The area between the cumulative distribution functions (ACDFs) of the model and observations quantifies the direction and magnitude of the distributional shift between the model and observations. It is calculated as:

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

$$\text{ACDFs} = \int_{-\infty}^{+\infty} [F_{model}(x) - F_{obs}(x)]dx \quad (2)$$

where F_{model} and F_{obs} are the CDFs of the model and observations, respectively. A positive (negative) value indicates that the model systematically underestimates (overestimates) the metric relative to observations.

2.4. Landfall Locations

We discretize global coastline into evenly spaced milepost points using the 1:110m Natural Earth shapefile (www.naturalearthdata.com). Mileposts are equally spaced along the coastline at approximately 1.4° arc-length intervals. For each storm track, the location that the TC first transits from ocean to land is identified as the landfall location. This point is then assigned to the nearest milepost based on geographic distance. The method is applied consistently to both observational and modelled tracks, enabling direct spatial comparison of landfall frequencies.

2.5. MTCE Identification

MTCEs are defined as two or more TCs that simultaneously coexist within the same basin (Fu et al., 2023; Schenkel, 2016). In this study, the algorithm for identifying MTCE clusters is implemented as follows: (1) The genesis and dissipation times of each individual TC track are merged into a single timeline and sorted chronologically. (2) An MTCE cluster is defined to commence when the active TC count reaches or exceeds two, persists for as long as this condition is maintained, and terminates once the count falls below two.

For each identified MTCE cluster, the maximum value of the active TC count during its duration is defined as the peak concurrent TC count for that cluster.

2.6. Return Period Calculation

The return period $RP(x)$ is defined as the average time interval between events (e.g., landfall intensity, inter-genesis time between MTCEs) whose magnitude exceeds a specified threshold x:

$$RP(x) = \frac{1}{[\lambda(1 - F(x))]} \quad (3)$$

where $F(x)$ is the cumulative distribution function and λ is the occurrence rate per unit time. In this study, for discrete observational or model data, $RP(x)$ is computed empirically as:

$$RP(x) = \frac{T_{total}}{N(x)} \quad (4)$$

where T_{total} is the total sampling period (i.e., the total observation period or the sum of the total simulation period across all model realizations), and $N(x)$ is the number of events exceeding threshold x.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

3 Model Evaluation

This section evaluates the performance of SPIN in reproducing key TC characteristics. Specifically, we assess annual frequency and seasonal cycle, patterns of TC genesis and track, and intensity statistics at the global scale. Special attention are paid on MTCEs metrics in the WP and NA basins, where such events are most frequent (Fu et al., 2025). For MTCE evaluation, model simulations are compared with observations and benchmarks introduced in section 2.3.

3.1 Inter-annual Variability and Seasonal Cycle

To evaluate the model's ability to represent interannual variability in TC activity, we calculated the correlation coefficients between basin-scale TC frequency in SPIN and observations over 1980-2022. As shown in Figure 1, correlation coefficients of all basins are statistically significant at the 0.05 level. Consistent with previous statistical downscaling models (Lee et al., 2018; Lin et al., 2023), SPIN performs best in the EP and NA basins ($r = 0.62$ and $r = 0.42$, respectively). The correlations may be further increased if the NeuralGCM simulations are initialized in April (instead of October in the previous year) and restricted evaluation to the TC season (June to December for Northern Hemisphere) (Figure S3). Moreover, SPIN demonstrates noticeably higher skills over previous statistical downscaling models in the WP and Southern Hemisphere basins (Lin et al. 2023). A possible explanation of this improvement is that in statistical downscaling frameworks (e.g., JL23), the interannual genesis signals can be smoothed in regions where ENSO variability has mixed-signed influences (Camargo et al., 2007; Lin et al., 2023). Specifically, the spatially random seeding approach used in their genesis component may artificially increase genesis probability in low-activity regions while decreasing it in high-activity regions, such that positive and negative biases across the basin may offset when summed, especially in mixed-signed ENSO signal regions. In contrast, SPIN extracts TCs directly from the environmental fields simulated by NeuralGCM, rather than relying on random seeding, thereby improving performance in regions with mixed ENSO signals. Overall, these results highlight the potential of SPIN to capture interannual variability in TC activities across most basins.

For the seasonal cycles of TC genesis (Figure 2), SPIN reproduces the bimodal distribution in the NI basin and the unimodal patterns in other basins, with a slight tendency to overestimate TC counts from the early to peak season and underestimate them from the peak to late season. A similar bias was noted by Zhang et al. (2025 , supplementary) that evaluates the performance of NeuralGCM.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

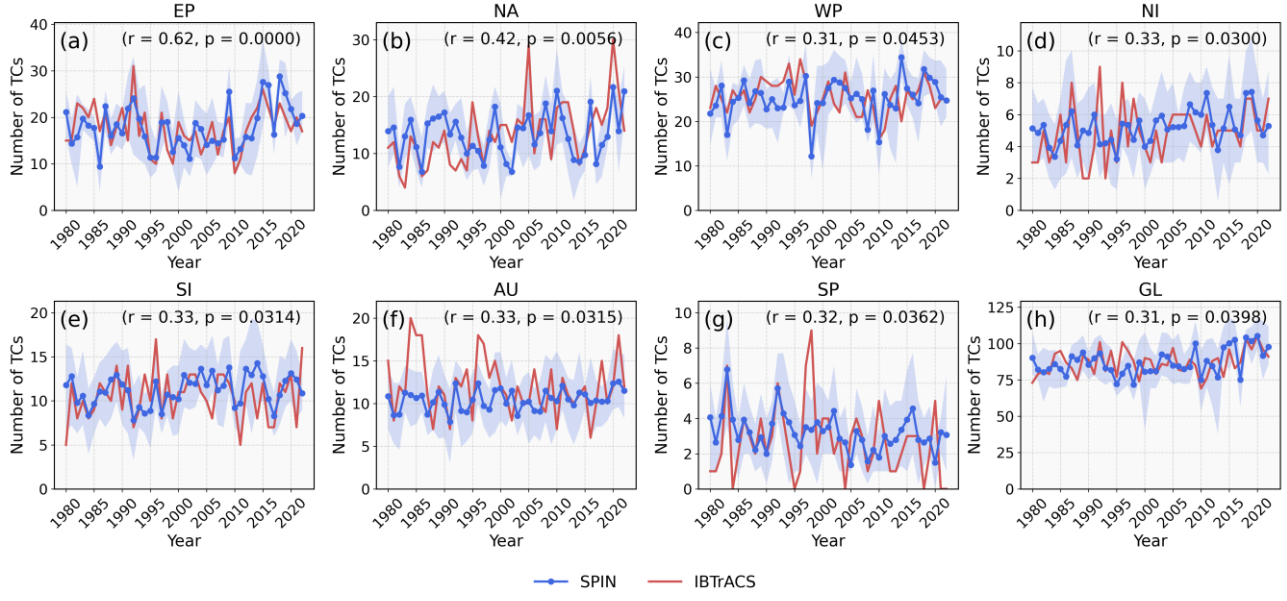
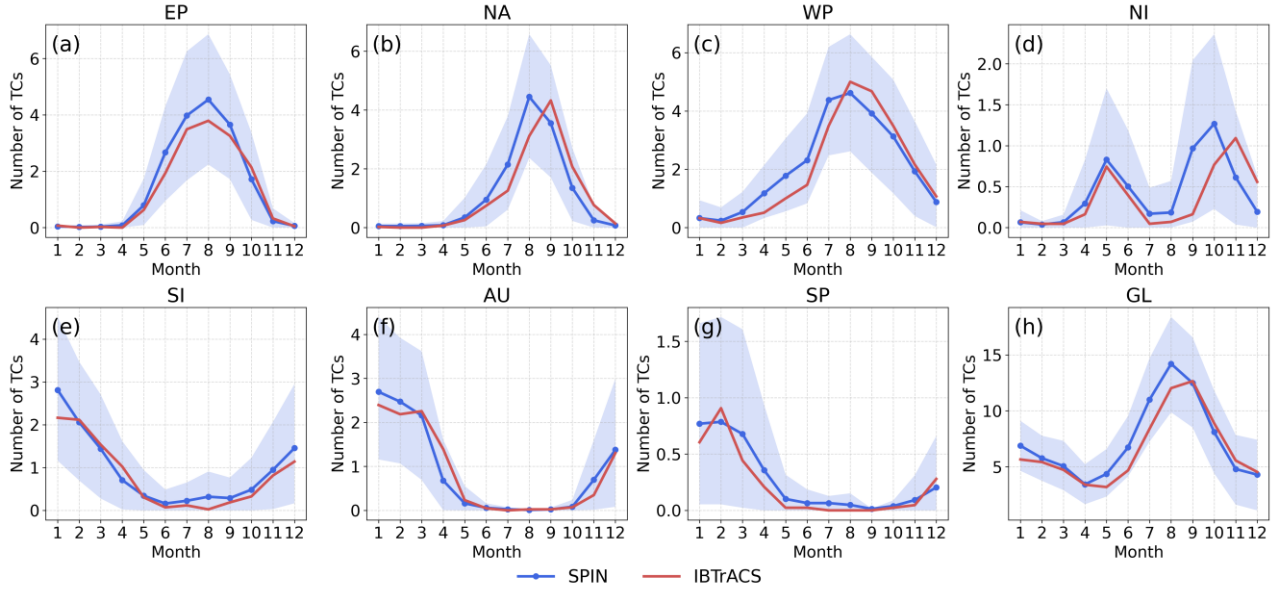


Figure 1. Interannual variability of TC frequency during 1980–2022. Panels show (a) EP, (b) NA, (c) WP, (d) NI, (e) SI, (f) AU, (g) SP, and (h) GL basins. The blue line shows the ensemble mean of 14-member SPIN simulations initialized in mid-October at 6-hour intervals, with shading indicating the 10th–90th percentile range, and observations are shown in red. Pearson correlation coefficients with significance levels are reported in the upper-right of each panel.



https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

Figure 2. Basin-wide seasonal cycle of TCs genesis averaged over 1980–2022. Panels show (a) EP, (b) NA, (c) WP, (d) NI, (e) SI, (f) AU, (g) SP, and (h) GL basins. The ensemble mean of SPIN simulations is shown in blue with shaded bands indicating the 10th-90th percentile range, and observations are shown in red. Only storms with lifetime maximum intensity exceeding 18 m s^{-1} are included.

3.2 Genesis and Track Statistics

We first compare the spatial pattern of SPIN-simulated TC genesis patterns with observations from 1980–2022. 3° × 3° grid boxes are used to calculate the genesis density (Figure 3). Overall, the spatial pattern of simulated genesis events aligns well with historical observations, though some regional biases remain. For example, SPIN tends to underestimate genesis frequency in the Bay of Bengal, the South China Sea, the Philippine Sea, the northern part of the AU basin, the eastern part of the EP basin, the Gulf of Mexico, and the western part of the NA basin. In contrast, it tends to overestimate genesis in the southern parts of the Eastern Pacific and NA basins, as well as in the northern part of the WP basin. Several reasons may contribute to this bias. First, TC numbers are sensitive to the detection scheme; although we have fine-tuned the detection method to allow TC numbers to align with observations at the basin scale, subregional differences remain. Second, in SPIN, NeuralGCM is forced with monthly SST and SIC fields linearly interpolated to 12-hour resolution. The interpolated forcing fields lack sub-monthly scale variability, which may also introduce bias. In addition, because genesis is defined as the first time the simulated surface wind of a TC exceeds 18 m s^{-1}, the bias in the intensity model may also influence the location of genesis events. Nevertheless, despite these biases, SPIN demonstrates skillful performance in reproducing the observed spatial distribution of TC genesis.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

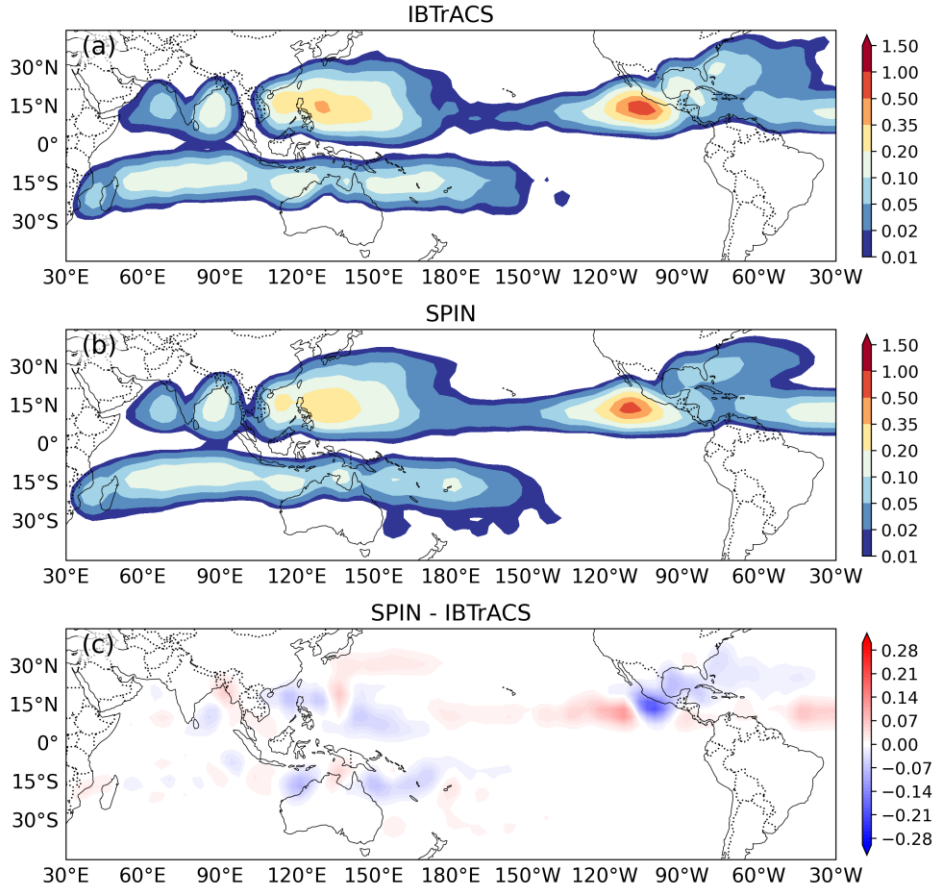


Figure 3. Tropical cyclones (TCs) genesis density per year on a 3° × 3° grid with Gaussian smoothing averaged over 1980–2022. (a) IBTrACS, (b) ensemble-mean simulations from SPIN, and (c) their difference (b − a). Genesis is defined as the first occurrence time that a TC's maximum wind speed ≥ 18 m s^{-1}.

To validate SPIN's performance in reproducing characteristics of TC movements, a comparison of the 6-hourly meridional and zonal displacements of simulated and observed tracks is shown in Figure 4. SPIN tends to underestimate the magnitude of the meridional displacement, suggesting that simulated biases in the steering flow (e.g., the subtropical high and trough structures) may appear overly smoothed (Duan et al., 2025). Such biases can in turn affect the recurvature latitude. Nevertheless, the overall simulation results show good agreement with observations.

Next, we validate the annual track density simulated by SPIN. We calculate the track density using 1° × 1° grid boxes (Figure 5). The spatial pattern of the bias of track density resembles that of genesis density, with SPIN underestimating density in the Bay of Bengal, southern part of WP basin, the northern part of the AU basin, the northern part of the EP and NA basin. These biases in track density are therefore largely attributable to bias in simulated genesis locations. Differences in TC duration

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

may also contribute to the bias in track density, with SPIN producing a mean global TC duration of 5.4 days versus 4.7 days in the observation.

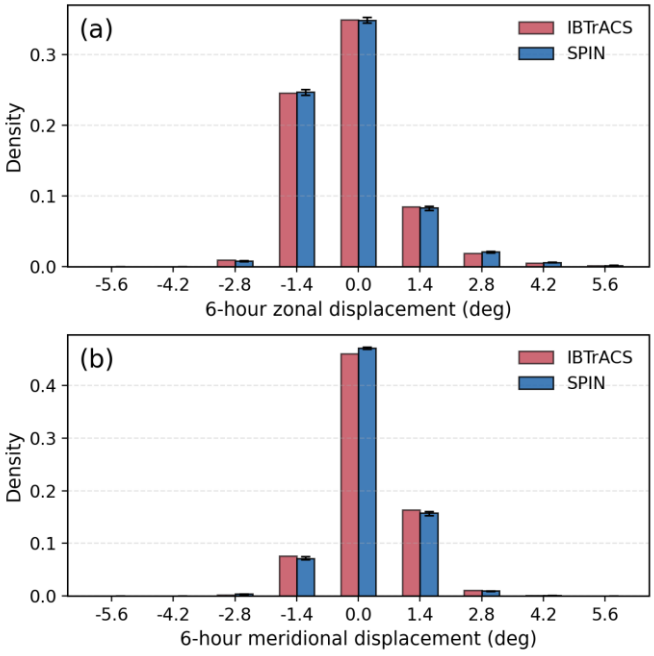


Figure 4. Comparison of six-hour TC displacement statistics between observations and the SPIN ensemble. (a) Zonal and (b) meridional displacements of TCs in IBTrACS (red) and SPIN (ensemble mean, blue). Error bars indicate the 10–90% ensemble spread for SPIN. Bins are defined at 1.4° intervals.

Next, we validate the annual track density simulated by SPIN. We calculate the track density using 1° × 1° grid boxes (Figure 5). The spatial pattern of the bias of track density resembles that of genesis density, with SPIN underestimating density in the Bay of Bengal, southern part of WP basin, the northern part of the AU basin, the northern part of the EP and NA basin. These biases in track density are therefore largely attributable to bias in simulated genesis locations. Differences in TC duration may also contribute to the bias in track density, with SPIN producing a mean global TC duration of 5.4 days versus 4.7 days in the observation.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

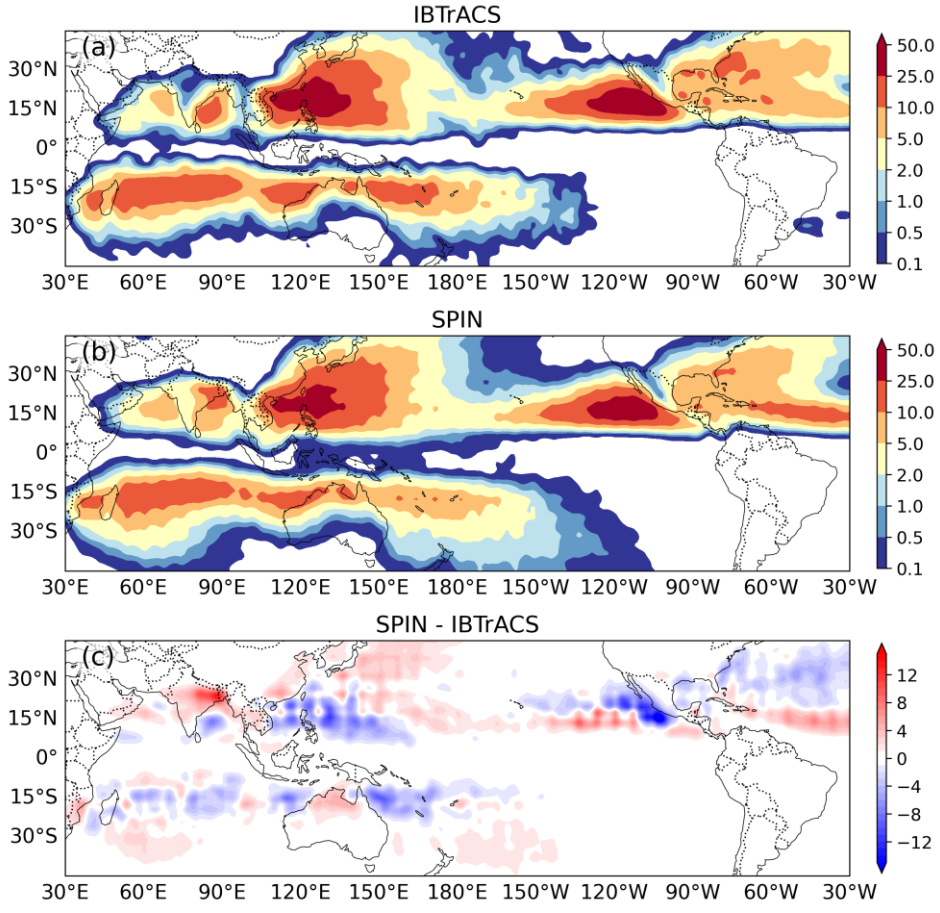


Figure 5. Annual 6-hourly TCs track density on a 1° × 1° grid with Gaussian smoothing for 1980–2022. (a) IBTrACS, (b) ensemble-mean simulations from SPIN, and (c) their difference (b − a). Only storms reaching maximum wind ≥ 18 m s^{-1} are included, with tracks evaluated from the first occurrence of that threshold.

3.3 Lifetime Maximum Intensity

We compare the SPIN-simulated lifetime maximum intensity (LMI) of TCs with observations at the global scale and in the basins with most TC-related economic losses, i.e., the NA and WP basins (Krichene et al., 2023) (Figure 6). In addition, the LMI derived from the JL23 model is included for comparison, which also adopts the FAST intensity model. This comparison allows us to determine whether the bias in SPIN simulated LMI originates from the FAST model or not.

SPIN generally well-reproduces the observed distribution of LMI, showing a peak followed by an exponential decay. This behavior is also captured by the JL23-downscaled ERA5 (hereafter JL-ERA5) and the JL23-downscaled NeuralGCM-simulated environments (hereafter JL-NGCM, representing the bias inherited from the large-scale environment simulated by NeuralGCM). The close agreement with observed distributions is supported by low $D_{\mathrm{KL}}(M \parallel O)$ values across basins (0.009-

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

0.029 in SPIN, 0.029-0.034 in JL-ERA5, 0.026-0.035 in JL-NGCM; Table S1). Both models exhibit a rightward shift of the peak, as indicated by consistently negative ACDF values (-0.853 to -3.353 in SPIN, -1.009 to -3.133 in JL-ERA5, -1.07 to -1.414 in JL-NGCM; Table S1), reflecting a modest overestimation of high-intensity events. In addition, both models underestimate the frequency of rapid intensification events (15.5-28.4% in SPIN, 14.5-26.6% in JL-ERA5, 16.9-28.1% in JL-NGCM, and 23.3-33.2% in observations across basins; Figure S4). These biases are therefore inherited from the FAST intensity model (Emanuel 2017). We also compared the LMI distributions from JL-ERA5 and JL-NGCM. The results show that JL-NGCM yields an LMI distribution similar to that derived from JL-ERA5, indicating that NeuralGCM simulated environmental fields carry negligible bias for TC intensity simulation.

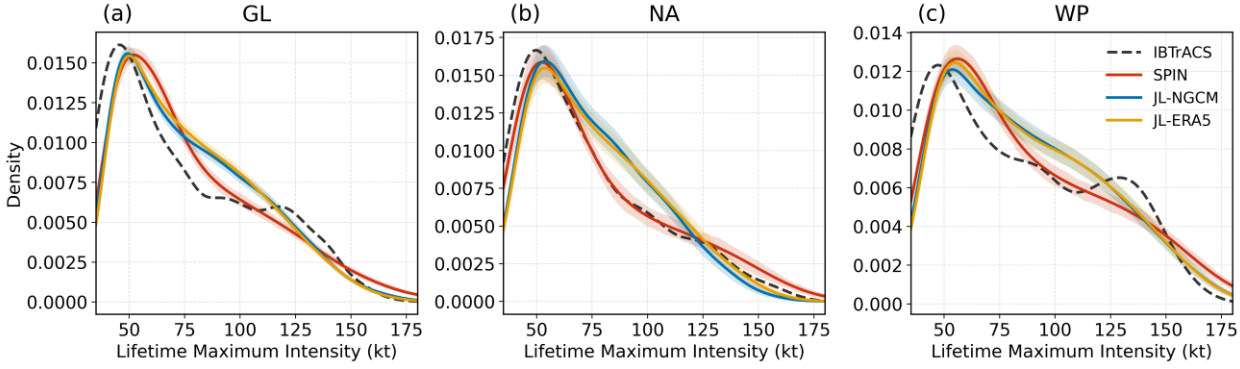
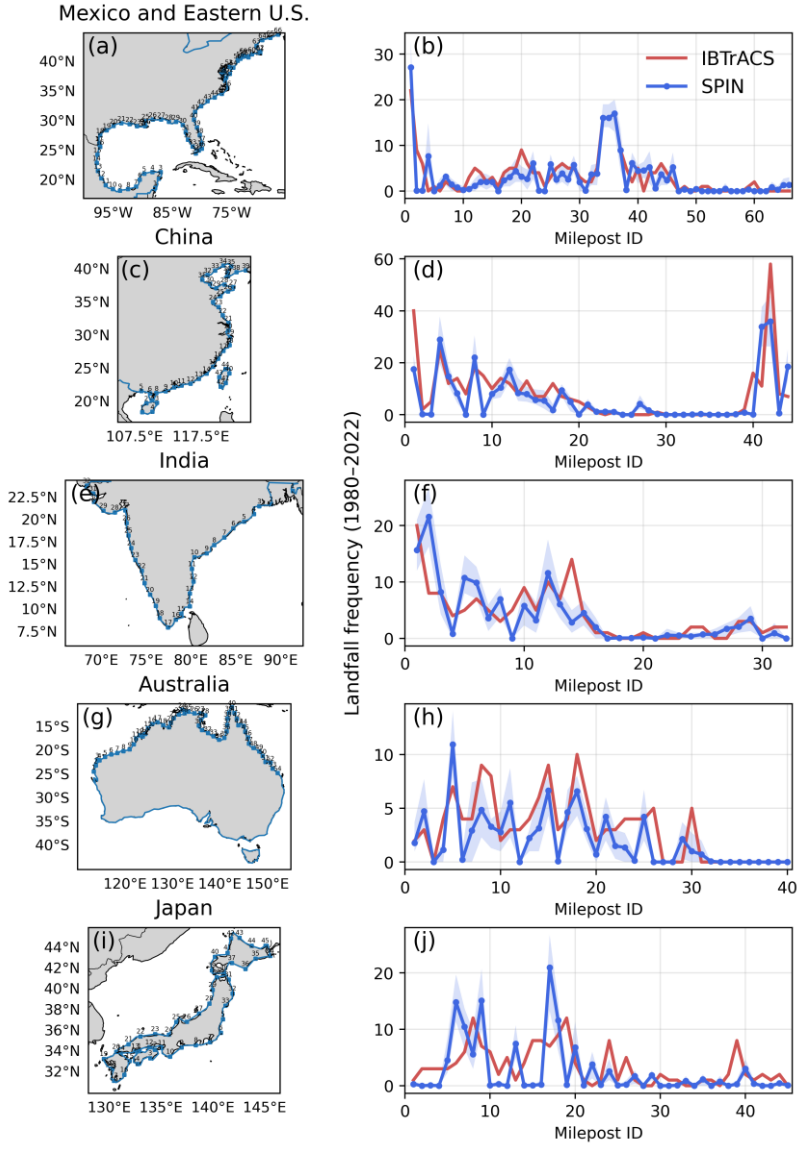


Figure 6. Distribution of LMI of TCs during 1980–2022. (a) shows the global distribution, (b) the NA basin, and (c) the WP basin distributions. Distributions are estimated using Gaussian kernel density fitting. Results of IBTrACS are shown as black dashed lines, SPIN simulations in red, JL-downscaled NeuralGCM (using a single ensemble member) in blue, and JL-downscaled ERA5 in yellow. For both JL-downscaled NeuralGCM and ERA5, the number of TCs is obtained by sub-sampling the downscaling events to match the number of observed events. The 10th-90th percentile ranges are indicated by shading.

3.4 Landfall frequency

To investigate the model's performance in landfalling TC characteristics, we evaluated the simulated landfall frequency across the main TC-impacted regions (Figure 7). The comparison shows SPIN has negligible bias relative to observations, particularly for Mexico, the eastern United States, and China. In India, Australia, and Japan, spatial patterns of the bias in landfall frequency largely mirror those in track density. Also, due to the complex geometry of the coastline and the coarser spatial resolution of the modeled tracks, the first ocean-to-land crossing of a modeled track may not correspond to the actual landfall location, thereby leading to incorrect milepost assignments. Despite these biases, SPIN successfully reproduces the spatial pattern of landfall hotspots.

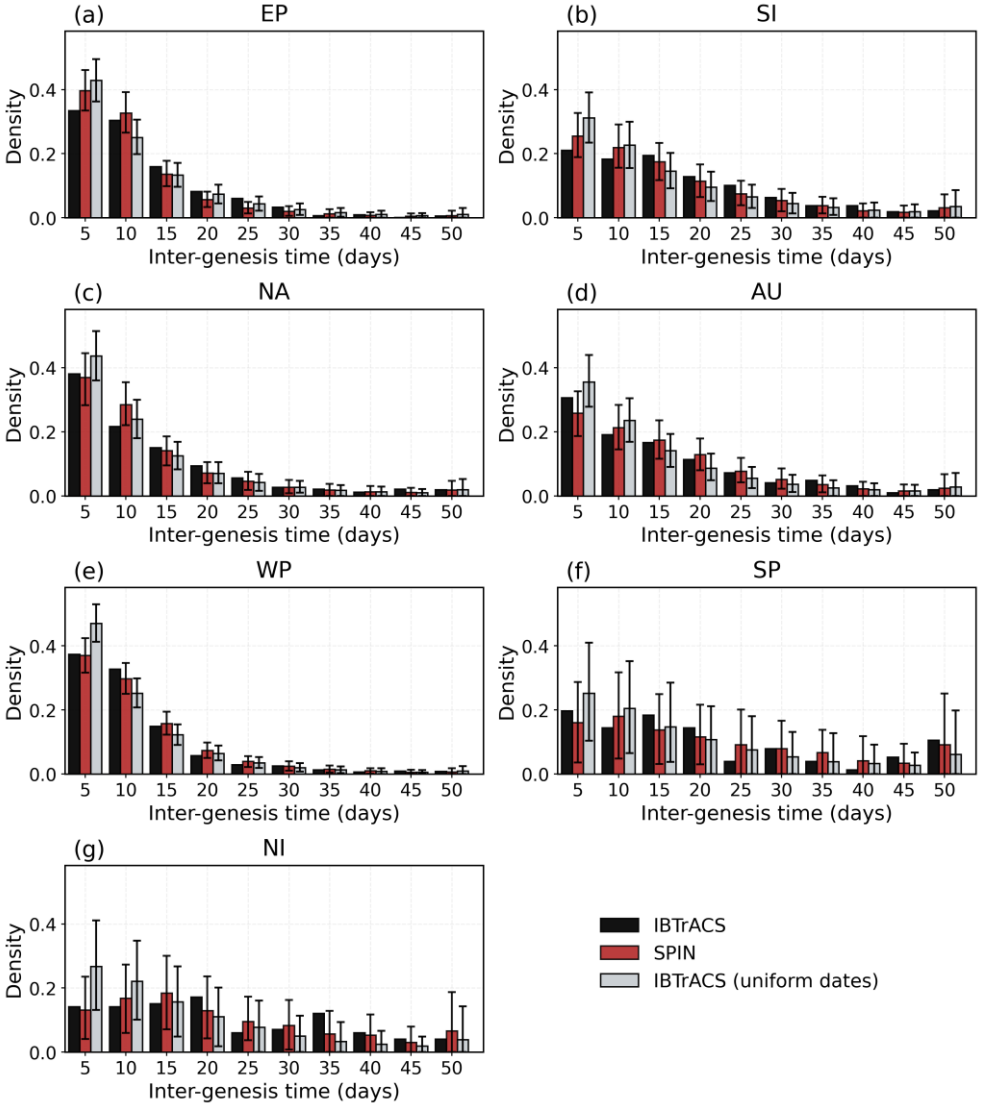
https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026



285

Figure 7. Landfall frequency of TCs along coastline mileposts during 1980-2022. Mileposts are equally spaced along the coastline at approximately 1.4° arc-length intervals for (a) Mexico and the eastern United States, (c) Eastern China, (e) India, (g) Australia, and (i) Japan. The SPIN ensemble-mean landfall frequencies are shown in (b), (d), (f), (h), and (j) in blue, with shaded bands indicating the 10th-90th percentile range across ensemble members. Observations are shown in red.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026



290

Figure 8. Histogram of TC inter-genesis dates from 1980-2022. Calculations are limited to the TC season in each year, defined as April–November for Northern Hemisphere basins (EP, NA, WP, NI; panels a, c, e, g) and from November to the following April for Southern Hemisphere basins (SI, AU, SP; panels b, d, f). Results from IBTrACS (black), IBTrACS with uniformly sampled dates (grey), the ensemble
295 mean of SPIN simulations (red) are shown for comparison. The 10th-90th percentile ranges are indicated in error bars.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

3.5 Inter-genesis dates

We assess whether SPIN realistically reproduces TC inter-genesis times in each basin (Figure 8). Because traditional statistical downscaling models (e.g., JL23) provide only monthly information on TC genesis, previous studies on back-to-back TCs derived event onset times by sampling from a Poisson–Gaussian model parameterized by seasonal variability (Xi et al., 2023; Xi and Lin, 2021). This procedure assumes that individual TCs are independent and thus lacks physical realism. Such challenge also extends to MTCEs, as pre-existing TCs may influence the genesis of subsequent ones (Ritchie and Holland, 1999; Yoshida and Ishikawa, 2013), violating the assumption of independence.

We include IBTrACS with uniformly sampled genesis dates (hereafter IBTrACS-uniform) for comparison, providing a clean reference of TC inter-genesis time distribution without the influence of biases from statistical downscaling models. SPIN exhibits lower $D_{\mathrm{KL}}(M \parallel O)$ values than IBTrACS-uniform in most basins, indicating a closer match to observations (0.007-0.062 for SPIN versus 0.019-0.168 for IBTrACS-uniform; Table S2). Additionally, IBTrACS-uniform tends to show a stronger left skew in its distribution, particularly within the 0-5 day bin, with ACDF values ranging from 0.25 to 0.63, compared to −0.24 to 0.31 for SPIN (Table S2). Given that the global average TC lifetime is around 5 days, overestimation of IBTrACS-uniform in 0-5 day may lead to an overestimation of MTCE hazards. Overall, these results highlight SPIN's improved capability in capturing the TCs inter-genesis dates, suggesting its potential to provide a more reliable basis for MTCE estimates.

3.6 Evaluation of MTCEs

In addition to evaluating individual TC events, this study also assesses MTCEs, a dimension rarely explored in earlier synthetic TC model evaluations. Specifically, we evaluate the performance of SPIN in simulating MTCEs in the WP and NA basins by examining MTCE inter-annual variability, the peak number of concurrent TCs per cluster, and the spatial relationship between pre-existing and subsequent TCs. For comparison, the JL23 model is also used to downscale both ERA5 and NeuralGCM-simulated large-scale environments. Like most statistical downscaling models, the JL23 model reports tracks on a monthly basis rather than giving precise genesis and dissipation dates. We therefore assign dates by uniformly sampling within each month for the tracks produced by the JL23 model, following previous studies (Xi et al., 2023; Xi and Lin, 2021).

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

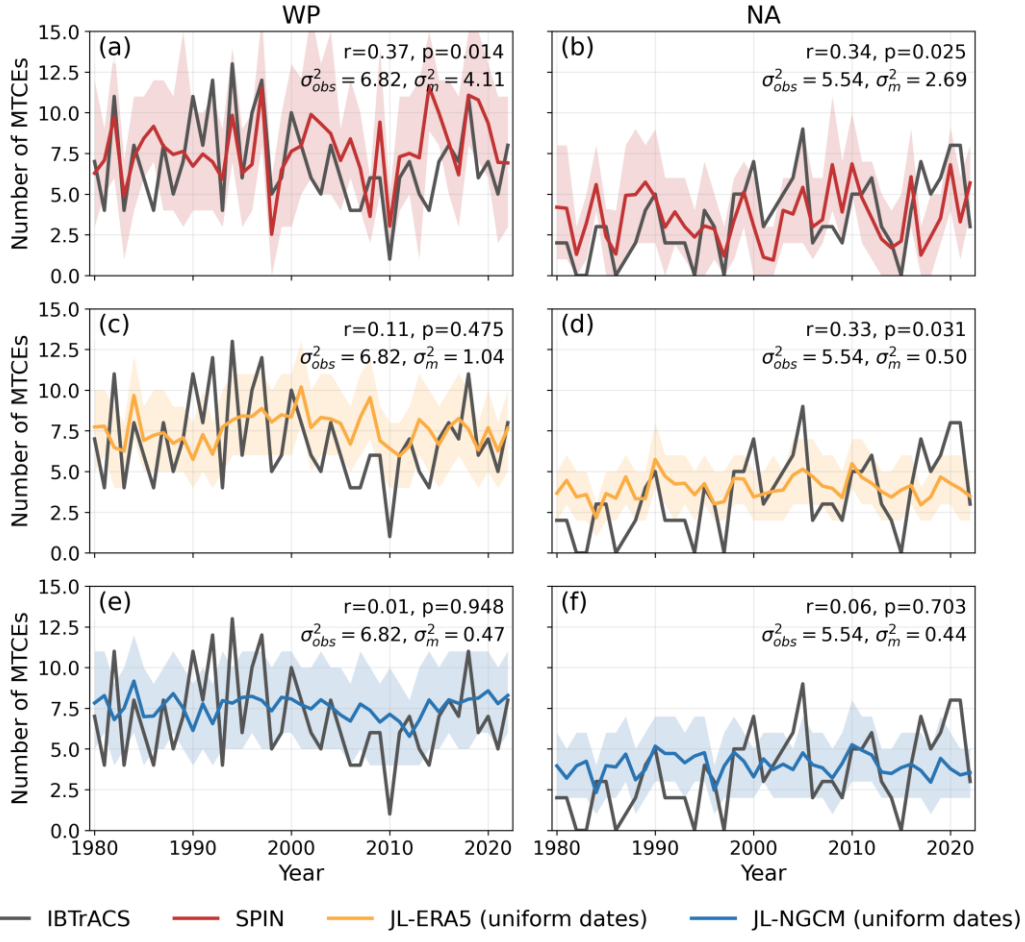


Figure 9. Interannual variability of multiple tropical cyclone events (MTCEs) in the WP and NA basins during 1980–2022. Results for IBTrACS are shown in black. The ensemble mean of SPIN simulations is shown in red (a-b), the JL23 model downscaling with uniformly sampled dates is shown in yellow for ERA5 (c-d) and in blue for NeuralGCM (e-f), with shaded bands indicating the 10th-90th percentile range. Pearson correlation coefficients with significance levels between the observations and the model are shown in the upper-right corner of each panel, along with the variances of the modelled $\boldsymbol{\sigma^2_m}$ and observed $\boldsymbol{\sigma^2_{obs}}$ annual frequencies.

Here, we begin by evaluating the interannual variability of MTCEs, which is inherently influenced by the interannual variability of the TC frequency. As shown in Figure 9, SPIN successfully reproduces the observed year-to-year fluctuations of MTCEs in both the WP and NA basins, with statistically significant correlations with observations. In contrast, JL23 performs poorly in the WP basin, where its simulated TC annual frequency shows little correlation with observations (Lin et al., 2023), leading to insufficient interannual variability of MTCEs. Moreover, SPIN simulated MTCEs better capture the observed variance, while JL23 downscaled results exhibit markedly lower variance, likely because the model synthesizes TCs from monthly mean environmental fields, which smooths out environmental conditions favoring MTCEs.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

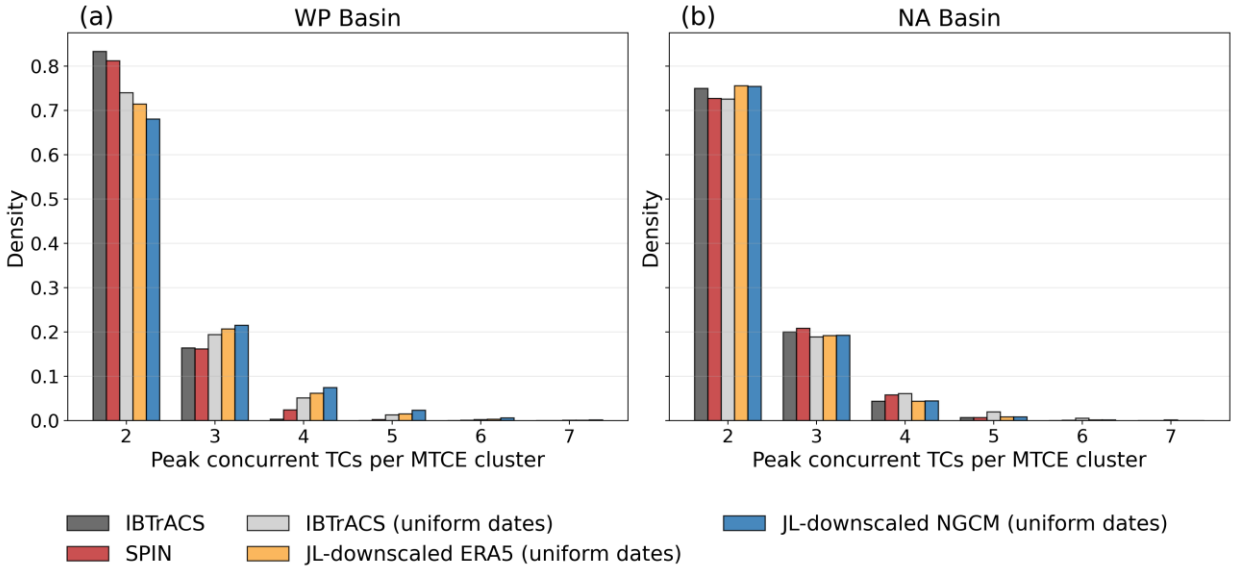


Figure 10. Histogram of MTCEs cluster size for the (a) WP and (b) NA basins during 1980–2022. The x-axis denotes the peak concurrent TC count per cluster, and the y-axis denotes the relative frequency. Results from IBTrACS (black), IBTrACS with uniformly sampled dates (grey), the ensemble mean of SPIN simulations (red), the JL downscaling of ERA5 with uniformly sampled dates (yellow), and the JL downscaling of NeuralGCM (using a single ensemble member) with uniformly sampled dates (blue) are shown for comparison.

Next, we evaluate the distribution of MTCE cluster sizes, defined as the peak number of concurrent TCs within a cluster (see section 2.5; Figure 10). SPIN closely reproduces the observed distribution, with MTCEs most frequently exhibiting two peak concurrent TCs. The probability decreases with increasing numbers of concurrent TCs per MTCE cluster, and clusters with more than five concurrent TCs are rare. In contrast, uniform-date sampling method, whether based on IBTrACS track-month statistics or applied to synthetic TCs simulated by JL23 model forced by NeuralGCM or ERA5, generally overestimates the frequency of larger clusters, especially in the WP basin. This bias may be linked to the overestimation of short inter-genesis times (0-5 days) for TCs by the uniform-date sampling method (Figure 8), which artificially increases the likelihood of multiple TCs being included within the same cluster. These results show that explicitly accounting for interactions between TCs and the environment in the downscaling model framework substantially improves realism for physically dependent co-occurring events.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

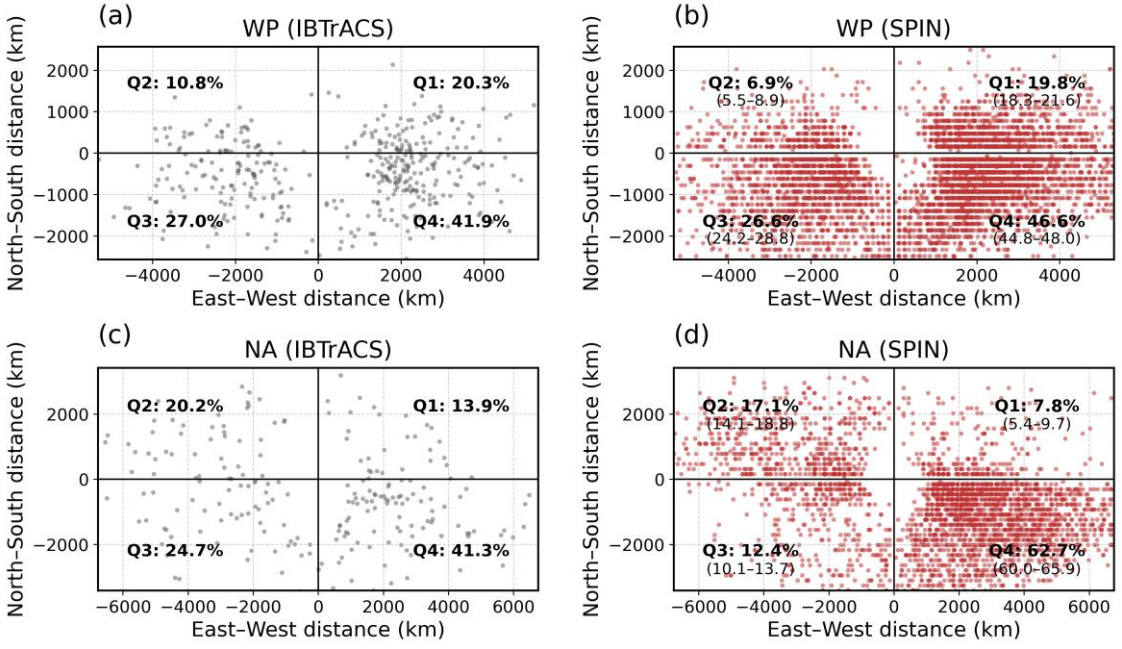


Figure 11. Relative locations between pre-existing and subsequent TCs in MTCEs from 1980 to 2022 for the WP (a–b) and NA (c–d) basins. The crosshair center marks the pre-existing TC. Points in the four quadrants show the relative positions of subsequent TCs for IBTrACS (first column) and for SPIN (second column; all ensemble members with a single velocity initialization run). Percentages for each quadrant are reported. SPIN results include the 10th-90th percentile range.

SPIN also well-captures the spatial pattern of MTCE compared to observation (Figure 11). SPIN reproduces the observed preference for subsequent TC genesis in the southeastern quadrant relative to the pre-existing TC (Krouse and Sobel, 2010; Yoshida and Ishikawa, 2013), with the fewest events occurring in the northwestern quadrant in the WP basin and in the northeastern quadrant in the NA basin. However, the percentage of events in each quadrant simulated by SPIN does not fully match the values observed. For example, SPIN overestimates the share of subsequent genesis events in southeastern quadrant in NA basin (Figure 3). For inter-storm distances (Figure 12), SPIN reproduces the observed distribution in both basins, with fewest events at <1000 km, a peak at 2000–3000 km in the WP, and a peak at >4000 km in the NA. SPIN slightly overestimates the >4000 km bin in both basins, which may be related to the overestimation of lifetime in the simulated TCs.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

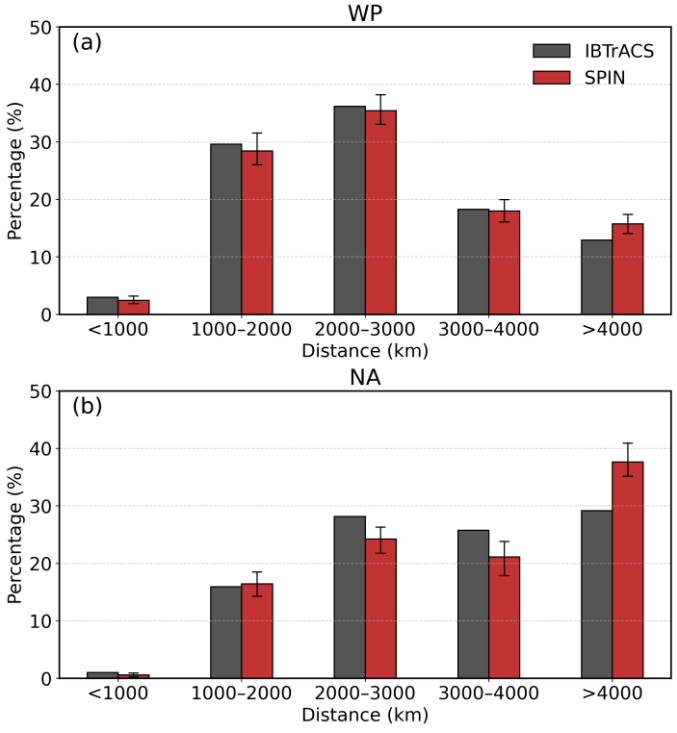


Figure 12. Histogram of distances between pre-existing and subsequent TCs within MTCEs from 1980 to 2022. Panel (a) shows the global distribution, panel (b) the NA basin, and panel (c) the WP basin distributions. Results are binned into fixed 1000-km intervals. IBTrACS are shown in black bar, and the SPIN ensemble mean is shown in red with 10th-90th percentile ranges.

4 Return Periods

In this section, we evaluate SPIN's skill in estimating return periods of metrics of interest at both global and regional scales. In particular, we investigated return periods of landfall intensity and the inter-event time between MTCEs.

4.1 Return period of landfall intensity

Figure 13 presents global maps of the return period for Category 1 and above TCs based on observations and SPIN downscaled results. SPIN successfully captures the main hazard hotspots, including the southwestern part of the western North Pacific, which covers southeastern China, southern Japan, and the Philippines, as well as the southeastern part of the eastern North Pacific. In these regions, return periods are generally less than 10 years and, in some areas, less than 5 years. Compared to observations, SPIN tends to overestimate hazard, indicated by shorter return periods, along the coasts of China, India, Australia, and the eastern United States; while it underestimates hazard, indicated by longer return periods, along the west coast of the United States. These differences are generally consistent with the overestimation of storm activity in these regions, as shown

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

in Figure 5. It should be noted that, since TCs are rare events and the observational record is limited in duration, the observed hazard may not necessarily reflect the true underlying risk (Lee et al., 2018; Lin et al., 2023).

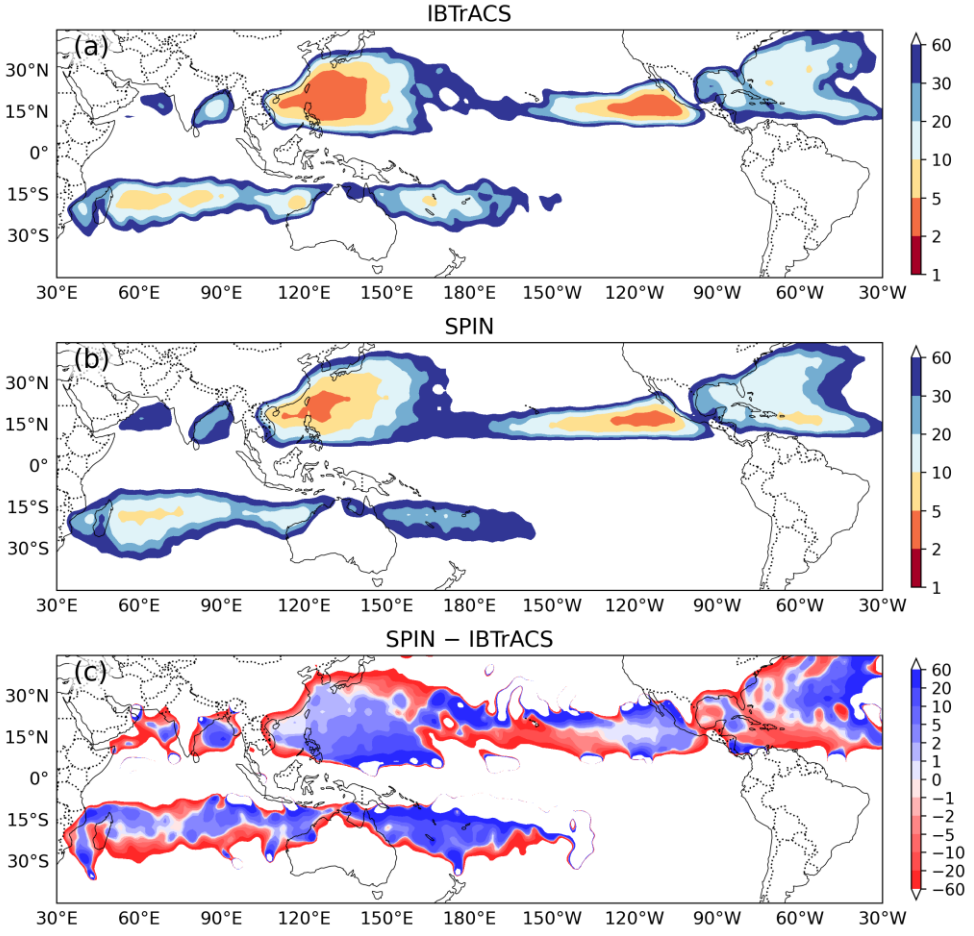


Figure 13. Global return period of TCs reaching ≥ 33 m s⁻¹ (Category 1): (a) IBTrACS, (b) SPIN ensemble mean, and (c) their difference on a 1°×1° grid with Gaussian smoothing. In panel (c), blue (red) indicates longer (shorter) return periods in SPIN than in IBTrACS, i.e., hazard underestimation (overestimation) by the model.

Next, we evaluate return period of landfall intensity for various TC-prone regions worldwide (Figure 14), with regional definitions following Lee et al. (2018) and Lin et al. (2023). The curves are presented without bias correction to the observations. Overall, the modelled return periods show good agreement with observations at lower intensities, although some biases remain, such as an overestimation of the return period for weak TCs in Gulf of Mexico, Philippines, and Madagascar. In contrast, the agreement is strong in the Eastern United States, China, and Japan. Biases in the return periods reflect biases in both landfall frequency and intensity. Further improvements may be achieved by tuning the intensity model and TCs tracking parameters.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

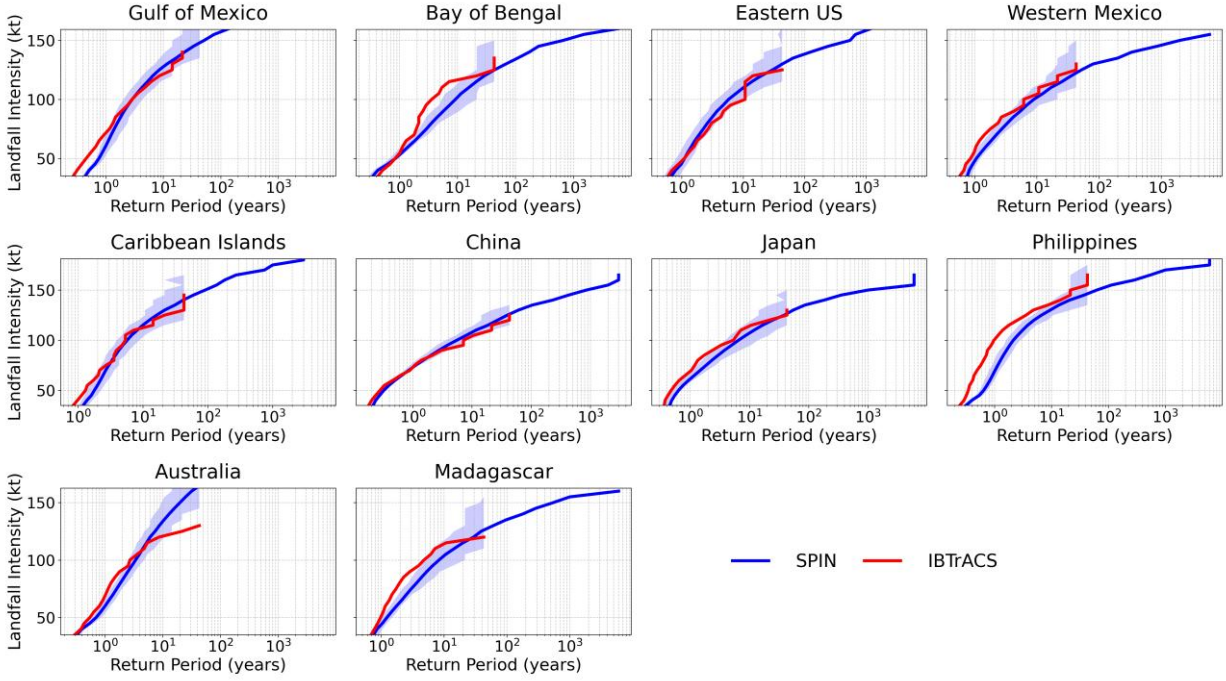


Figure 14. Return periods of landfall intensity for the following regions: Gulf of Mexico, Bay of Bengal, Eastern United States, Western Mexico, Caribbean Islands, China, Japan, Philippines, Australia, and Madagascar. Results from IBTrACS are shown in red, the solid blue line shows the return period curve computed from all ensemble members of SPIN combined (i.e., stacking all 14 members, 43 years, and 10 runs, yielding a total of 6,020 years). The shaded region represents the 10th–90th percentile spread across all individual ensemble members.

4.2 Return period of MTCEs

Finally, we present multi-model simulated return periods for the inter-MTCE time between MTCEs, as shown in Figure 15. SPIN closely matches the observed return period in both basins. In contrast, JL-downscaled ERA5 and JL-downscaled NeuralGCM tend to underestimate the interval, likely related to their insufficient ability to represent MTCE. IBTrACS-uniform matches the observed return periods in both basins, but cannot provide estimates for the tail of the curve due to sampling limitations. Moreover, because this approach does not simulate TCs based on environmental fields, it is difficult to extend to future climate scenarios. Whether differences between these return periods will increase or decrease under warming, and what physical mechanisms control the change, remain important and open questions for future research.

Overall, compared to previous synthetic TC models, SPIN achieves a more comprehensive coupling between environmental fields and TCs. This enables physically consistent MTCE return period estimates even with limited computational resources, providing a novel tool for assessing TC-related compound hazards.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

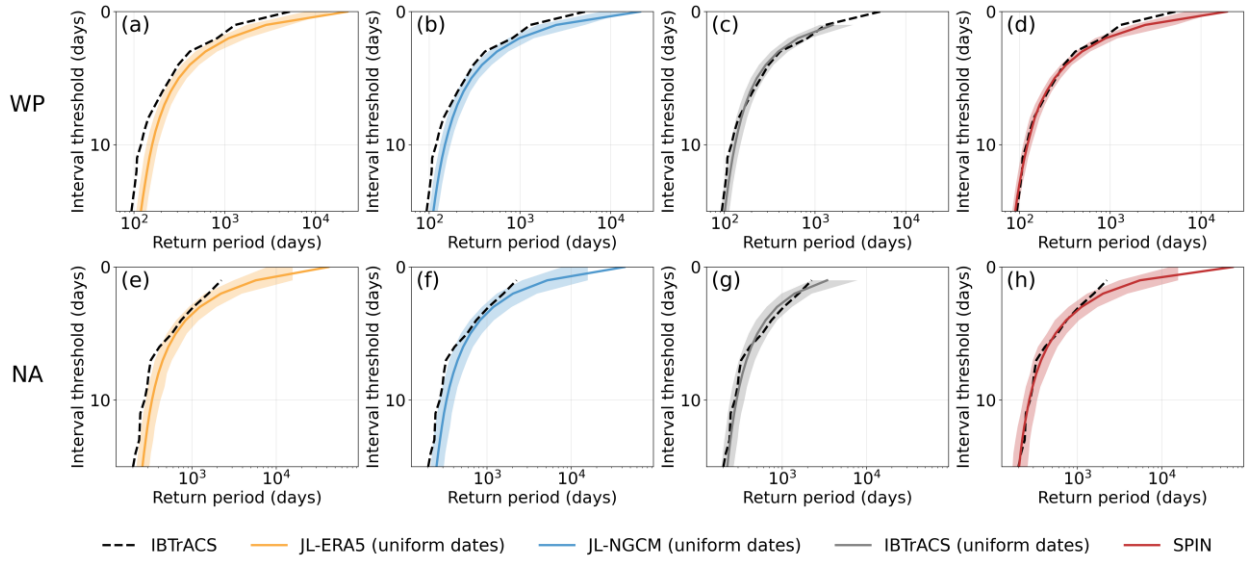


Figure 15. Return periods of inter-MTCE time. The y-axis shows the MTCE inter-genesis interval (0–15 days), and the x-axis shows the return period (days). Panels (a–d) correspond to the WP basin and panels (e–h) to the NA basin. Shown for comparison are results from IBTrACS (black), IBTrACS with uniformly sampled dates (grey), JL-downscaled ERA5 with uniformly sampled dates (yellow), and JL-downscaled NeuralGCM (single ensemble member) with uniformly sampled dates (blue), and SPIN (red). The solid line in each panel indicates the return period computed from all Monte Carlo simulations (a, b, c, e, f, g) or all ensemble members (d, h). The shaded region represents the 10th–90th percentile range across individual ensemble members.

5 Summary and Discussion

In this study, we develop a hybrid framework, SPIN, that functions as a spontaneous synthetic TC model with realistic intensity, to downscale TCs for risk analysis. The model leverages NeuralGCM (Kochkov et al., 2024) to simulate ensembles of spontaneously generated TC tracks, and couples the FAST intensity model (Emanuel, 2017; Lin et al., 2023) to compute realistic intensities based on the large-scale environment. In this way, SPIN bridges the gap between statistical and dynamical downscaling models, enabling the generation of large ensembles of physically consistent synthetic TCs at low computational cost. In contrast to traditional statistical downscaling models that assume independent TCs, SPIN explicitly simulates TC emergence from and interactions with the environment, enabling improved representation of multiple TC events (MTCEs).

SPIN reproduces the observed climatology of TC activity reasonably, including interannual variability of TC frequency, seasonal cycle, patterns of TC genesis and tracks, and the probability distribution of LMI. Compared with previous statistical downscaling models, SPIN better represents interannual variability in regions where ENSO signals (e.g. the WP and Southern Hemisphere basins) are mixed. This advantage of SPIN stems from simulated TCs interacting with the environment, whereas random seeding methods applied in previous synthetic TC models can artificially enhance genesis in low-TC-activity regions while suppressing it in high-TC-activity regions (Camargo et al., 2007; Lin et al., 2023). Furthermore, unlike statistical models

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

that provide only monthly information on TC genesis, SPIN explicitly outputs the TC genesis and lysis time, thus can more realistically reproduce the distribution of inter-TC-genesis dates in each basin.

Beyond reproducing individual TC events, SPIN shows robust skills in simulating MTCEs. We evaluate SPIN's performance in reproducing MTCE characteristics in the WP and NA basins, where MTCEs occur most frequently (Fu et al., 2025). SPIN reproduces the interannual variability of MTCEs more faithfully than the benchmark models, exhibiting higher correlations with observations and a closer match in interannual variance. SPIN reproduces the peak number of concurrent TCs within MTCE clusters more accurately, benefiting from its improved representation of TC inter-genesis timing. In addition, it captures the spatial distribution and return periods of MTCEs with close agreement with observations.

While SPIN provides a physically consistent and computationally efficient framework for synthetic TC simulation, several aspects can be further improved in future works. For example, although the FAST intensity model component within SPIN reproduces the overall distribution of LMI reasonably well, it tends to underestimate the occurrence of TC rapid intensification and in turn resulted in fewer occurrence of extreme intense TCs compared with observation. Moreover, since SPIN is driven by environmental fields simulated by NeuralGCM, biases can be introduced through the representation of large-scale climate features in NeuralGCM. Leveraging the modular design of our model, which allows seamless integration with other AI-based weather and climate models, SPIN can be further improved when future AI models generate more realistic large-scale climate features. In addition, the SPIN framework can be extended to explore other types of compound hazards, such as TC-heatwave interactions. It might also be coupled with GCM simulations to assess projected changes in TC and MTCE characteristics under future climate scenarios.

Code and data availability

The full code for the SPIN (v1.0) model, together with documentation and sample data is archived on Zenodo at https://doi.org/10.5281/zenodo.17863785 (Gao and Xi, 2025). The SPIN-simulated tracks for all ensembles analyzed in this study are archived on Zenodo under the copyright license CC BY 4.0 at https://doi.org/10.5281/zenodo.17863932 (Gao, 2025). The TempestExtremes tool (Ullrich et al., 2021) for tropical cyclone detection and characterization is available at https://doi.org/10.5281/zenodo.4385656. The snapshots of the JL23 benchmark model (Lin et al., 2023) and IBTrACS dataset (Knapp et al., 2010) used in this study are archived on Zenodo at https://doi.org/10.5281/zenodo.18230393 to ensure reproducibility. All original credit for the benchmark model and dataset belongs to their respective authors.

Author contributions

DX conceived the study. YG developed the model and carried out the simulations. Both authors discussed and interpreted the results. YG drafted the manuscript, and DX contributed to its revision.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

Competing interests

The authors declare there are no conflicts of interest in this manuscript.

Acknowledgments

The computations were performed using the research computing facilities provided by the Information Technology Services at the University of Hong Kong. Both authors acknowledge support from the HKU Start-up Fund (000250348.130087.25300.100.01) and the New Staff Seed Fund (103032009.376). We also thank Paul Ullrich for valuable guidance on the use of TempestExtremes package.

References

Baxter, I., Pahlavan, H., Hassanzadeh, P., Rucker, K., and Shaw, T.: Benchmarking atmospheric circulation variability in an AI emulator, ACE2, and a hybrid model, NeuralGCM, https://doi.org/10.48550/arXiv.2510.04466, 6 October 2025.

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

Bloemendaal, N., Haigh, I. D., Moel, H. de, Muis, S., Haarsma, R. J., and Aerts, J. C. J. H.: Generation of a global synthetic tropical cyclone hazard dataset using STORM, Sci. Data, 7, 40, https://doi.org/10.1038/s41597-020-0381-2, 2020.

Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Allen, A., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J. A., Dong, H., Gupta, J. K., Thambiratnam, K., Archibald, A. T., Wu, C.-C., Heider, E., Welling, M., Turner, R. E., and Perdikaris, P.: A foundation model for the Earth system, Nature, 641, 1180–1187, https://doi.org/10.1038/s41586-025-09005-y, 2025.

Buonomo, E., Savage, N., Dong, G., Becker, B., Jones, R. G., Tian, Z., and Sun, L.: Tropical Cyclone Changes in Convection-Permitting Regional Climate Projections: A Study Over the Shanghai Region, J. Geophys. Res. Atmospheres, 129, e2023JD038508, https://doi.org/10.1029/2023JD038508, 2024.

Camargo, S. J., Emanuel, K. A., and Sobel, A. H.: Use of a Genesis Potential Index to Diagnose ENSO Effects on Tropical Cyclone Genesis, https://doi.org/10.1175/JCLI4282.1, 2007.

Davis, C. A.: Resolving Tropical Cyclone Intensity in Models, Geophys. Res. Lett., 45, 2082–2087, https://doi.org/10.1002/2017GL076966, 2018.

Duan, S., Zhang, J., Bonfils, C., and Pallotta, G.: Could NeuralGCM effectively simulate future heatwaves? A case study based on the 2021 Pacific Northwest heatwave, https://doi.org/10.48550/arXiv.2410.09120, 6 April 2025.

Emanuel, K.: Climate and Tropical Cyclone Activity: A New Model Downscaling Approach, https://doi.org/10.1175/JCLI3908.1, 2006.

Emanuel, K.: A fast intensity simulator for tropical cyclone risk analysis, Nat. Hazards, 88, 779–796, https://doi.org/10.1007/s11069-017-2890-7, 2017.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

Emanuel, K. and Zhang, F.: The Role of Inner-Core Moisture in Tropical Cyclone Predictability and Practical Forecast Skill, https://doi.org/10.1175/JAS-D-17-0008.1, 2017.

Emanuel, K., Ravela, S., Vivant, E., and Risi, C.: A Statistical Deterministic Approach to Hurricane Risk Assessment, Bull. Am. Meteorol. Soc., 87, 299–314, https://doi.org/10.1175/BAMS-87-3-299, 2006.

Fu, Z.-H., Zhan, R., Zhao, J., Yamada, Y., and Song, K.: Future Projections of Multiple Tropical Cyclone Events in the Northern Hemisphere in the CMIP6-HighResMIP Models, Geophys. Res. Lett., 50, e2023GL103064, https://doi.org/10.1029/2023GL103064, 2023.

Fu, Z.-H., Xi, D., Xie, S.-P., Zhou, W., Lin, N., Zhao, J., Wang, X., and Chan, J. C. L.: Shifting hotspot of tropical cyclone clusters in a warming climate, Nat. Clim. Change, 15, 850–858, https://doi.org/10.1038/s41558-025-02397-9, 2025.

Gao, Y.: SPIN (v1.0) outputs for the study "A Spontaneous Synthetic Tropical Cyclone Model Empowered by NeuralGCM for Hazard Assessment," Zenodo [dataset], https://doi.org/10.5281/zenodo.17863932, 2025.

Gao, Y. and Xi, D.: A hybrid framework for simulating spontaneous synthetic tropical cyclones (TCs) with realistic intensity: SPIN (v1.0), Zenodo [dataset], https://doi.org/10.5281/zenodo.17863785, 2025.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Jing, R. and Lin, N.: An Environment-Dependent Probabilistic Tropical Cyclone Model, J. Adv. Model. Earth Syst., 12, e2019MS001975, https://doi.org/10.1029/2019MS001975, 2020.

Jing, R., Gao, J., Cai, Y., Xi, D., Zhang, Y., Fu, Y., Emanuel, K., Diffenbaugh, N. S., and Bendavid, E.: TC-GEN: Data-Driven Tropical Cyclone Downscaling Using Machine Learning-Based High-Resolution Weather Model, J. Adv. Model. Earth Syst., 16, e2023MS004203, https://doi.org/10.1029/2023MS004203, 2024.

Judt, F., Klocke, D., Rios-Berrios, R., Vanniere, B., Ziemen, F., Auger, L., Biercamp, J., Bretherton, C., Chen, X., Düben, P., Hohenegger, C., Khairoutdinov, M., Kodama, C., Kornblueh, L., Lin, S.-J., Nakano, M., Neumann, P., Putman, W., Röber, N., Roberts, M., Satoh, M., Shibuya, R., Stevens, B., Vidale, P. L., Wedi, N., and Zhou, L.: Tropical Cyclones in Global Storm-Resolving Models, J. Meteorol. Soc. Jpn. Ser II, 99, 579–602, https://doi.org/10.2151/jmsj.2021-029, 2021.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J.: The International Best Track Archive for Climate Stewardship (IBTrACS), https://doi.org/10.1175/2009BAMS2755.1, 2010.

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., and Hoyer, S.: Neural General Circulation Models for Weather and Climate, Nature, 632, 1060–1066, https://doi.org/10.1038/s41586-024-07744-y, 2024.

Krichene, H., Vogt, T., Piontek, F., Geiger, T., Schötz, C., and Otto, C.: The social costs of tropical cyclones, Nat. Commun., 14, 7294, https://doi.org/10.1038/s41467-023-43114-4, 2023.

Krouse, K. D. and Sobel, A. H.: An observational study of multiple tropical cyclone events in the western north Pacific, Tellus Dyn. Meteorol. Oceanogr., 62, 256–265, https://doi.org/10.1111/j.1600-0870.2009.00435.x, 2010.

https://doi.org/10.5194/egusphere-2025-5540
Preprint. Discussion started: 14 January 2026

Lee, C., Tippett, M. K., Sobel, A. H., and Camargo, S. J.: An Environmentally Forced Tropical Cyclone Hazard Model, J. Adv. Model. Earth Syst., 10, 223–241, https://doi.org/10.1002/2017MS001186, 2018.

Lin, J., Emanuel, K., and Vigh, J. L.: Forecasts of Hurricanes Using Large-Ensemble Outputs, https://doi.org/10.1175/WAF-D-19-0255.1, 2020.

Lin, J., Rousseau-Rizzi, R., Lee, C.-Y., and Sobel, A.: An Open-Source, Physics-Based, Tropical Cyclone Downscaling Model With Intensity-Dependent Steering, J. Adv. Model. Earth Syst., 15, e2023MS003686, https://doi.org/10.1029/2023MS003686, 2023.

Price, I., Sanchez-Gonzalez, A., Alet, F., Andersson, T. R., El-Kadi, A., Masters, D., Ewalds, T., Stott, J., Mohamed, S., Battaglia, P., Lam, R., and Willson, M.: Probabilistic weather forecasting with machine learning, Nature, 637, 84–90, https://doi.org/10.1038/s41586-024-08252-9, 2025.

Ritchie, E. A. and Holland, G. J.: Large-Scale Patterns Associated with Tropical Cyclogenesis in the Western Pacific, 1999.

Schenkel, B. A.: A Climatology of Multiple Tropical Cyclone Events, https://doi.org/10.1175/JCLI-D-15-0048.1, 2016.

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., and Reed, K. A.: TempestExtremes v2.1: a community framework for feature detection, tracking, and analysis in large datasets, Geosci. Model Dev., 14, 5023–5048, https://doi.org/10.5194/gmd-14-5023-2021, 2021.

Xi, D. and Lin, N.: Sequential Landfall of Tropical Cyclones in the United States: From Historical Records to Climate Projections, Geophys. Res. Lett., 48, e2021GL094826, https://doi.org/10.1029/2021GL094826, 2021.

Xi, D., Lin, N., and Gori, A.: Increasing sequential tropical cyclone hazards along the US East and Gulf coasts, Nat. Clim. Change, 13, 258–265, https://doi.org/10.1038/s41558-023-01595-7, 2023.

Yoshida, R. and Ishikawa, H.: Environmental Factors Contributing to Tropical Cyclone Genesis over the Western North Pacific, https://doi.org/10.1175/MWR-D-11-00309.1, 2013.

Zhang, G., Rao, M., Yuval, J., and Zhao, M.: Advancing seasonal prediction of tropical cyclone activity with a hybrid AI-physics climate model, Environ. Res. Lett., 20, 094031, https://doi.org/10.1088/1748-9326/adf864, 2025.

Zhou, Y. and Lin, Y.: A Toy Model for the Global Annual Number of Tropical Cyclones, Earths Future, 12, e2024EF004839, https://doi.org/10.1029/2024EF004839, 2024.

Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., and Zhang, X.: Future climate risk from compound events, Nat. Clim. Change, 8, 469–477, https://doi.org/10.1038/s41558-018-0156-3, 2018.