






RESEARCH ARTICLE

10.1029/2023MS004203

TC-GEN: Data-Driven Tropical Cyclone Downscaling Using Machine Learning-Based High-Resolution Weather Model

Renzhi Jing^{1,2} , Jianxiang Gao³, Yunuo Cai⁴, Dazhi Xi⁵ , Yinda Zhang⁶, Yanwei Fu⁴, Kerry Emanuel⁷, Noah S. Diffenbaugh^{1,8,9} , and Eran Bendavid^{2,10}

Special Collection:

Machine learning application to Earth system modeling

Key Points:

- We introduce a novel approach to tropical cyclone (TC) downscaling using a machine learning based global weather model, named “TC-GEN”
- We generate and integrate synthetic TC seeds into the surrounding environment using a data-driven approach
- TC-GEN establishes a framework that opens the possibility of modeling the two-way interactions between storms and the environment

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

R. Jing and Y. Fu,
rjing@stanford.edu;
yanweifu@fudan.edu.cn

Citation:

Jing, R., Gao, J., Cai, Y., Xi, D., Zhang, Y., Fu, Y., et al. (2024). TC-GEN: Data-driven tropical cyclone downscaling using machine learning-based high-resolution weather model. *Journal of Advances in Modeling Earth Systems*, 16, e2023MS004203. <https://doi.org/10.1029/2023MS004203>

Received 27 DEC 2023

Accepted 10 SEP 2024

Author Contributions:

Conceptualization: Renzhi Jing, Noah S. Diffenbaugh, Eran Bendavid
Data curation: Renzhi Jing
Formal analysis: Renzhi Jing, Jianxiang Gao, Yanwei Fu, Kerry Emanuel

© 2024 The Author(s). Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](https://creativecommons.org/licenses/by/4.0/), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

¹Woods Institute for the Environment, Stanford University, Stanford, CA, USA, ²School of Medicine, Stanford University, Stanford, CA, USA, ³Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai, China, ⁴School of Data Science and MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China, ⁵Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA, ⁶Google LLC, Mountain View, CA, USA, ⁷Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA, ⁸Department of Earth System Science, Stanford University, Stanford, CA, USA, ⁹Doerr School of Sustainability, Stanford University, Stanford, CA, USA, ¹⁰Department of Health Policy, Stanford University, Stanford, CA, USA

Abstract Synthetic downscaling of tropical cyclones (TCs) is critically important to estimate the long-term hazard of rare high-impact storm events. Existing downscaling approaches rely on statistical or statistical-deterministic models that are capable of generating large samples of synthetic storms with characteristics similar to observed storms. However, these models do not capture the complex two-way interactions between a storm and its environment. In addition, these approaches either necessitate a separate TC size model to simulate storm size or involve post-processing to capture the asymmetries in the simulated surface wind. In this study, we present an innovative data-driven approach for TC synthetic downscaling. Using a machine learning-based high-resolution global weather model (ML-GWM), our approach can simulate the full life cycle of a storm with asymmetric surface wind that accounts for the two-way interactions between the storm and its environment. This approach consists of multiple components: a data-driven model for generating synthetic TC seeds, a blending method that seamlessly integrates storm seeds into the surrounding while maintaining the seed structure, and a model based on a recurrent neural network to correct for biases in storm intensity. Compared to observations and synthetic storms simulated using existing statistical-deterministic and statistical downscaling approaches, our method shows the ability to effectively capture many aspects of TC statistics, including track density, landfall frequency, landfall intensity, and outermost wind extent. Leveraging the computational efficiency of ML-GWM, our approach shows substantial potential for TC regional hazard and risk assessment.

Plain Language Summary Tropical cyclones (TCs) cause significant destruction each year. It is crucial to accurately assess the risks they present, but this is challenging due to a scarcity of historical data. A commonly used approach involves creating a large number of synthetic TCs that share key characteristics with real storms. However, traditional synthetic TC generation approaches do not capture the complex interactions between storms and their larger-scale environment. Furthermore, these approaches do not adequately represent the asymmetric structure of TCs. Recently, advances in machine learning-based global weather model (ML-GWM) have provided highly accurate and efficient high-resolution global weather forecasts that surpass conventional numerical weather forecasting. Here, we introduce a novel synthetic TC generation approach, which we call the synthetic TC-GENerative Model (or “TC-GEN”), leveraging the state-of-the-art ML-GWM. We show that TC-GEN can generate a large number of synthetic storms that allow the interaction between the storm and its environment. We evaluate the performance of TC-GEN in various aspects, including several landfall characteristics, which are of the most importance for local TC risk analysis. Our study also serves as a compelling example of the transformative impact of machine learning and data science in revolutionizing climate studies during the era of artificial intelligence.

1. Introduction

Tropical cyclones (TCs) are among the most destructive natural disasters, causing substantial damage and losses in multiple ocean basins annually. In a warming climate, it is projected that TCs are likely to become more intense, with an expected increase in both the maximum peak wind speed and the proportion of strong TCs in the future (H. Lee et al., 2023; Pörtner et al., 2022). Accurate assessment of TC tracks and intensities is fundamental

Funding acquisition: Renzhi Jing, Yanwei Fu, Noah S. Diffenbaugh, Eran Bendavid
Investigation: Renzhi Jing
Methodology: Renzhi Jing, Jianxiong Gao, Yunuo Cai, Dazhi Xi, Yinda Zhang, Kerry Emanuel, Noah S. Diffenbaugh, Eran Bendavid
Project administration: Renzhi Jing, Yanwei Fu, Eran Bendavid
Resources: Yanwei Fu
Software: Jianxiong Gao, Yinda Zhang
Supervision: Renzhi Jing, Yanwei Fu, Kerry Emanuel, Noah S. Diffenbaugh, Eran Bendavid
Validation: Renzhi Jing, Jianxiong Gao
Visualization: Renzhi Jing, Jianxiong Gao
Writing – original draft: Renzhi Jing
Writing – review & editing: Renzhi Jing, Jianxiong Gao, Yinda Zhang, Kerry Emanuel, Noah S. Diffenbaugh, Eran Bendavid

to reducing the impacts of landfalling storms. However, with around 90 storms occurring every year and an average of only 20 making landfall, this task is challenging due to the shortage of historical data required for regional risk assessment. To overcome data deficiency, a widely used approach is to generate synthetic TCs that are capable of responding to various climate conditions (Emanuel et al., 2008; Jing & Lin, 2020; C.-Y. Lee et al., 2018). Using large samples of synthetic storms, including extreme events with extended return periods, enables a comprehensive risk assessment for specific regions in both current and future climates.

Previous studies on synthetic TC generation have primarily employed two main approaches: (a) statistical re-sampling, and (b) physical-based downscaling methods. Statistical re-sampling models simulate TC genesis, tracks, and intensities (maximum wind speed or minimum central pressure) based purely on historical observational data sets, without considering environmental conditions. Examples of such models include those developed by Vickery et al. (2000), James and Mason (2005), Bloemendaal et al. (2020). These models typically require a limited number of input variables, have low computational costs, and thus are easily applicable on a global scale. However, these models are not based on physical principles and cannot be accurately applied to a non-stationary climate due to changes in the background environment. On the other hand, physically based downscaling methods relate TC characteristics to the large-scale background environmental conditions, making them environment-dependent. These methods can be statistical-deterministic, (e.g., models developed by Emanuel et al., 2006) or purely statistical (e.g., models developed by Jing & Lin, 2020; C.-Y. Lee et al., 2018). Such environment-dependent approaches are capable of simulating the TC climatology in future climate scenarios and therefore are suitable for climate change studies (Emanuel et al., 2008; Jing et al., 2021; C.-Y. Lee et al., 2020). Since the first synthetic TC downscaling approach in this family of models appeared in 2006, significant advances have been made in each of the three components (Huang et al., 2021; Huang, Wang, Jing, et al., 2022), and these approaches have been widely used for applications such as TC-induced surge risk assessment (Lin et al., 2012), regional loss assessment (Huang, Wang, Liu, et al., 2022; Meiler et al., 2022), and TC-induced wind load analysis (Kareem et al., 2019).

Both statistical and statistical-deterministic synthetic downscaling methods simulate the complete lifecycle of TCs using environmental parameters derived from the background environment. However, in such approaches, the environment lacks the capability to respond to the evolution of the storm, and thus is unable to simultaneously simulate the two-way interactions between the storm and its surrounding environment. Furthermore, traditional approaches do not comprehensively simulate the characteristics of TCs (i.e., genesis, track, intensity, and size) as a cohesive system. For example, although the TC intensity is determined based on environmental predictors along the track, existing storm track models are predominantly driven by background winds, with storm intensity playing a lesser role. Given the evident and significant correlations, it is prudent for downscaling models to capture the potential interdependencies between these components (Ruan & Wu, 2022). Moreover, the size of the outermost wind extent of the TC and the asymmetries in the surface wind field are not directly captured. Some approaches do not output the TC size and require a separate size component to determine the outer radius of the storm by random sampling from historical data (Jing & Lin, 2020). Other methods provide the radius of maximum wind (Emanuel et al., 2008); however, they require an additional parametric wind model to generate the full surface wind field, followed by post-processing to incorporate asymmetries related to storm translation speed and wind shear (Lin & Chavas, 2012; Lin et al., 2012).

The ideal synthetic downscaling method would simultaneously simulate all characteristics of the TC as an integrated system, including the interactions between storms and the surrounding environment, to generate detailed wind fields with greater accuracy but similar computational efficiency to that of traditional statistical and statistical-deterministic downscaling methods. Recent progress in machine learning-based high-resolution global weather modeling (ML-GWM) (Bi et al., 2023a; Lam et al., 2022; Pathak et al., 2022) has made this possible. ML-GWM systems are based on three-dimensional neural networks that are trained on high-quality reanalysis data sets, such as the ECMWF reanalysis v5 (ERA5) (Hersbach et al., 2020), to predict weather around the globe. A significant advantage of ML-GWMs is their substantially lower computational costs compared to traditional numerical weather forecasting, while still operating at high spatial resolutions. Representing the cutting edge of ML-GWM, Pangu-Weather (Bi et al., 2023a) has outperformed the operational Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts in medium-range forecasting, with speeds more than 10,000 times faster. The high spatial resolution of 0.25° also enables Pangu-Weather to precisely track TCs based on simulation results.

In this study, we leverage ML-GWM to create a novel ML-based approach for synthetic TCs downscaling, which we call the synthetic TC-GENerative Model (or “TC-GEN”). This involves generating a synthetic TC seed for each storm through a data-driven process, merging it with the background environment, and simulating both the storm and its surroundings simultaneously with Pangu-Weather. To achieve this, we first determine the annual frequency, date, and location of synthetic TCs using the genesis model of PepC, an environment-dependent TC downscaling model (Jing & Lin, 2020). Next, we perform a principal component analysis (PCA) on all historical TCs at genesis, identifying the principal components that effectively capture most of the variances in TC genesis. Using these principal components, we generate synthetic TC seeds with weights derived from historical data (when referring to “TC seeds,” we mean early stage storms that have just reached genesis). We then integrate these TC seeds into the surrounding environment using Poisson blending, a technique widely used in image processing to seamlessly merge two images, ensuring that the TC seeds are naturally embedded within the larger environmental context while still maintaining important wind structures. Finally, we run Pangu-Weather using their pre-trained model, which enables the joint simulation of both the storm and its surrounding environment, and bias-correct simulated intensity to real intensity. This integrated approach allows us to gather key characteristics of the TC, such as the track and the maximum wind speed. It also provides spatial details such as the full wind field, allowing for a direct derivation of the outermost extent of the storm. It is worth noting that several key steps of this ML-based method are data-driven, relying heavily on historical data that lack substantial input from physics. Furthermore, while the spatial resolution of 0.25° is a high resolution for global climate models, it is still too coarse to accurately resolve the storm structure, as the effective resolution of machine learning-based weather forecasting models can be even coarse (Bonavita, 2023). Therefore, we should use the simulated three-dimensional structure including the horizontal surface wind field with care, as it is likely to be unrealistic.

Despite these limitations, our ML-based method offers a unique set of collective advantages compared to previous TC downscaling approaches: (a) Holistic simulation: Unlike previous studies in which the genesis, track, and intensity of the storm are simulated separately, this approach can simulate these three storm components together as a cohesive system; (b) Integrated simulation: Similar to numerical modeling, this approach simulate both the storm and its environment simultaneously, thus allowing for the two-way interactions between the storm and the environment; (c) Intrinsic asymmetry: While the TC core and intensity may not be fully resolved, this approach has the capability to provide crucial asymmetric characteristics in the TC surface wind field. This, in turn, allows for the inference of the asymmetric outermost wind extent of a storm, which is essential for studying TC-induced hazards such as surges and heavy rainfall; (d) Efficiency: Utilizing pre-trained Pangu-Weather, this approach inherits the efficiency of statistical downscaling methods that require little computational resources, enabling large samples of synthetic TCs to be generated in days, a time frame comparable to the work of Bloemendaal et al. (2020); (e) Scalability: This approach can be readily expanded to include other ocean basins. Moreover, the entire modeling framework is potentially applicable for TC downscaling using alternative high-resolution climate data sets, such as those available in HighResMIP (Haarsma et al., 2016). However, for the application of this approach, the climate data sets must possess sufficiently high resolution to effectively detect storm low-pressure center and offer a reasonable representation of the storm's outer size. Additionally, it is necessary to pre-train the ML-WNP using the pertinent climate data sets as training data to get a corresponding simulator core.

To evaluate TC-GEN, we generate a large sample of synthetic TCs and compare those simulated storms with historical observations. We also place TC-GEN in context with an existing statistical-deterministic downscaling approach, represented by Emanuel et al. (2008) (KE08) and a statistical downscaling method, represented by the Princeton environment-dependent probabilistic model (PepC, Jing & Lin, 2020). We choose these two existing approaches as they are both environment dependent and have distinct model components. The metrics we use for comparison include the density of the tracks over the ocean, the maximum lifetime intensity, the landfall frequency, and the landfall intensity distributions. Given that KE08 and PepC lack the ability to simulate the outer size of TC, we only compare the distribution of TC outer size simulated by TC-GEN with that of the historical TCs identified using reanalysis data. In all of these metrics, we demonstrate a strong alignment between simulated storms and observational data. We further assess the adaptability of TC-GEN to different reanalysis data sets through two analyses: one with the ERA5 reanalysis, on which Pangu-Weather is trained, but using different temporal resolutions for model initialization, and the other using an alternative reanalysis data set from the National Centers for Environmental Prediction (NCEP). We show that the effectiveness of TC-GEN depends on the consistency between the training reanalysis data set used to train the ML-GMW and the data used for model

initialization. Based on these analyses, we summarize the strength and limitations of TC-GEN and propose potential improvements for future work.

2. Pre-Trained Models and Data

2.1. Machine Learning-Based Global Weather Forecasting Model

In this study, we use Pangu-Weather as the core ML-GWM to generate synthetic TCs. Pangu-weather is an artificial intelligence-based model for medium-range global weather forecasting, which has been shown to outperform the operational IFS of the European Center for Medium-Range Weather Forecasts (ECMWF) (Bi et al., 2023a).

Pangu-Weather is trained on atmospheric reanalysis data from the ERA5 reanalysis data (Hersbach et al., 2020), using 69 atmospheric and surface variables as input and forecasting these variables for the subsequent time step. The 69 variables include 65 upper-air variables (geopotential height, specific humidity, temperature, u and v component of wind at 13 pressure levels) plus four surface weather variables (2 m temperature, u- and v-component of 10 m wind speed, and mean sea level pressure). Pangu-Weather offers four pre-trained models that are capable of forecasting global weather data 1, 3, 6, and 24 hr ahead. We use the model with a lead time of 6 hr since it is consistent with the temporal resolution of TC IBTrACS data (see Section 2.2). Pangu-Weather has the ability to produce accurate simulations that span multiple consecutive days, a time frame that is adequate to simulate the entire life span of the most TCs. Pangu-Weather is open-source, and the pre-trained models can be downloaded at <https://github.com/198808xc/Pangu-Weather>.

Pangu-Weather has several distinct advantages, making it highly suitable for this study. First, Pangu-Weather is among the state-of-the-art ML-GWM systems with the highest performance in weather forecasting. Second, Pangu-Weather has a high horizontal resolution of $0.25^\circ \times 0.25^\circ$. Although the thermodynamic processes of the TCs cannot be fully resolved at this spatial resolution, it is sufficient to detect the low pressure center, which enables effective storm eye tracking. Third, Pangu-Weather is computationally efficient (more than 10,000 times faster than operational dynamical models), making it capable of simulating large numbers of synthetic storms. In our experiments, we run Pangu-Weather with 8 RTX A6000 GPUs, which adequately support all our computational tasks.

2.2. Tropical Cyclone Data

TC observations are extracted from IBTrACS data (version v04r00) (Knapp et al., 2010, 2018), which includes the location (latitude and longitude) of each storm every 6 hr, along with its maximum sustained wind speed and minimum central pressure measured at a height of 10 m above the sea surface. In this study, historical TCs in the North Atlantic Basin are utilized for multiple purposes: (a) the time and location of historical TCs between 1979 and 2022 are used to extract TC genesis from the ERA5 reanalysis data; (b) the time and location of historical TCs between 1979 and 2022 are used to crop the ERA5 and Pangu data set; these data, along with the intensity evolutions of TCs are used to develop the intensity bias correction model; (c) the tracks and intensities of historical TCs between 1979 and 2022 are used to evaluate the performance of TC-GEN, including the track density, lifetime maximum intensity (LMI), and landfall intensity; and (d) extended TC data dating back to 1900 are used to analyze landfall frequency, which provides a more precise assessment of sampling errors with a larger number of TC tracks.

2.3. Reanalysis Data

We use reanalysis data with different temporal resolution and from multiple sources in this study to initiate Pangu-Weather, to develop the intensity bias correction model, and to evaluate the performance of synthetic storms generated by TC-GEN.

2.3.1. Pangu-Weather Initialization

Pangu-Weather requires global atmospheric and surface variables at the current time step to forecast the next step. To be consistent with the input setup for the Pangu-Weather model, we use ERA5 reanalysis with a resolution of $0.25^\circ \times 0.25^\circ$, the highest available spatial resolution, and a temporal resolution of both 6-hr and monthly reanalysis data to initiate Pangu-Weather. Using 6-hr reanalysis data for Pangu-Weather initialization is intuitive, as it

ensures alignment with the model setup and prevents domain gaps. However, generating synthetic TCs with 6-hourly reanalysis data presents two challenges. First, introducing a synthetic TC to the 6-hourly reanalysis data may result in multiple storms developing on the same map. This does not reflect reality, as the simultaneous occurrence of multiple storms is relatively rare (Chowdhury et al., 2022). Second, generating synthetic TCs by initiating Pangu-Weather with 6-hourly data requires downloading extensive historical ERA5 reanalysis data, resulting in substantial data storage requirements. In contrast, starting Pangu-Weather with monthly data significantly reduces the data storage burden and mitigates the problem of simultaneous occurrence. In this study, we use both 6-hourly and monthly reanalysis data to initialize Pangu-Weather, and under both scenarios we use the pre-trained model with a lead time of 6 hr to simulate storms. We denote these two scenarios as “TC-GEN hourly” and “TC-GEN monthly” in the following text and results. It is worth noting that in “TC-GEN monthly,” the Pangu-Weather still runs on a 6-hr basis. The key distinction between “TC-GEN hourly” and “TC-GEN monthly” lies in the data utilized for initialization. For instance, in simulating a synthetic storm that originates in August 2020, “TC-GEN monthly” is initialized with monthly mean reanalysis data, whereas “TC-GEN hourly” is initialized with 6-hourly reanalysis data.

In addition to ERA5 reanalysis, we test TC-GEN performance when initiating Pangu-Weather with the NCEP Global Forecast System (GFS) from Global Forecast Grids Historical Archive at $0.25^\circ \times 0.25^\circ$ resolution (NCEP et al., 2015). We opt for this reanalysis data set because it provides all 69 variables required by Pangu-Weather, and it is available at a resolution of $0.25^\circ \times 0.25^\circ$ with global coverage.

2.3.2. TC Seed Generation

To develop the model for synthetic TC seed generation, we extract atmospheric and surface fields that represent the seeds of historical TCs from the ERA5 6-hr reanalysis. These extracted fields are used for the PCA analysis. They are also used to evaluate the performance of the PCA model by comparing real TC seeds with reconstructed TC seeds with different bases. More details can be found in Section 3.2.

2.3.3. TC Intensity Bias Correction

To develop the TC intensity bias correction model, which adjusts maximum wind speed and minimum central pressure based on TC and environmental characteristics, we crop environmental fields of TC intensity predictors from the ERA5 6-hourly reanalysis data. These data are used to develop both the feature encoder and the main intensity evolution model. More details on model structure and model development can be found in Section 3.5.

2.3.4. TC Outer Size Evaluation

To assess the performance of the outer sizes of synthetic TCs simulated by TC-GEN, we use the outer sizes of historical TCs as the reference, which are represented in the ERA5 6 hr reanalysis. More details on the definitions of the TC outer sizes and how the outer sizes are computed from the ERA5 reanalysis data can be found in Section 3.6.

2.4. Existing Downscaling Approaches

We contextualize TC-GEN by comparing it with two existing synthetic downscaling approaches: the statistical-deterministic model developed by Emanuel et al. (2008) (KE08), and PepC, a statistical synthetic downscaling approach introduced in Jing and Lin (2020). Both synthetic downscaling approaches have three components: a genesis model, a track model, and an intensity model, but with different model specifications.

KE08 applies a random seeding method to initiate the storm, a beta and advection model based on local winds to propagate the storm, and a deterministic Coupled Hurricane Intensity Prediction System (CHIPS; Emanuel et al., 2004) model to estimate the intensity of the storm based on the local thermodynamic state of the atmosphere and ocean. The model has been extensively applied to assess TC hazards (Emanuel, 2017; Marsooli et al., 2019), economic losses (Meiler et al., 2022; Mendelsohn et al., 2012), and changes in TC climatology under projected future climate conditions (Emanuel et al., 2008; Jing et al., 2021). Here, we use a total of 4,100 synthetic TCs downscaled from ERA5 reanalysis during the period 1980–2020, which intensify and reach TC strength (LMI exceeds 34 kt).

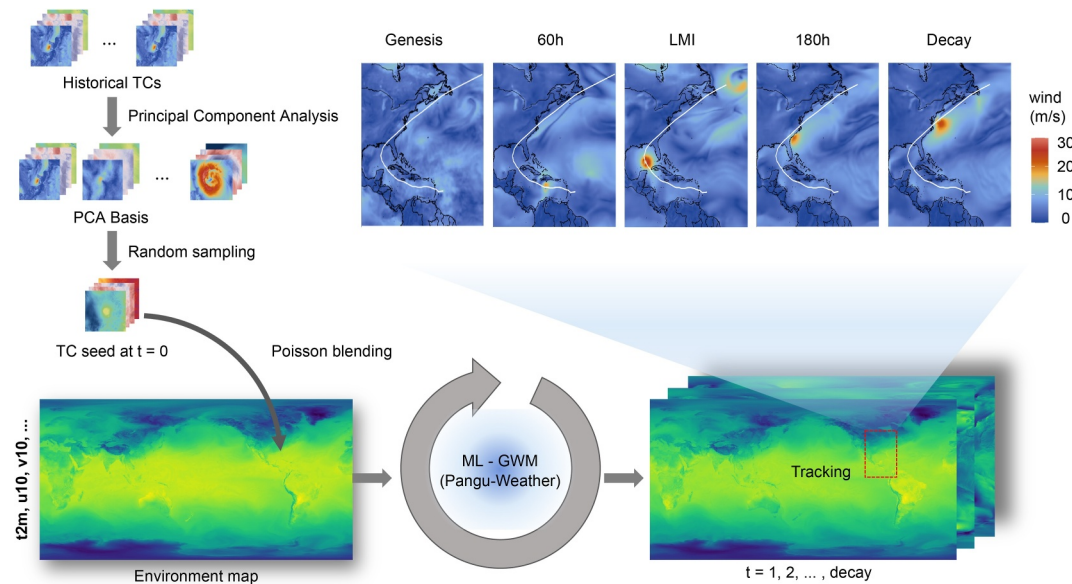


Figure 1. TC-GEN Workflow. We propose an AI-empowered approach for synthetic tropical cyclone (TC) downscaling where TC seeds are created through a data-driven approach using Principal Component Analysis, and subsequently integrated into the reanalysis environment map using Poisson blending. Then, both the TC and its surrounding environment are simulated in an interactive manner, using a high-resolution machine learning-based global weather model (Pangu-Weather in our study). The TC location is then tracked from the simulation output. In the upper right corner, we show the simulated track of a single synthetic TC as an illustration. The color field represents the total surface wind of the synthetic storm at five stages across its life cycle: genesis, 60 hr after genesis, the time when the storm reaches its maximum lifetime intensity, 180 hr after genesis, and finally, the decay stage.

In PepC, the genesis component determines annual frequency, as well as the time and location of weak vortices. The genesis model is developed using Poisson regression based on four large-scale environmental variables: absolute vorticity, wind shear, relative humidity, and maximum potential intensity (Emanuel, 1988), which are contained in a genesis index (Tippett et al., 2011). After initialization, an analog-wind track model is used to propagate the storm based on analog factors (from historical track patterns) and local in situ winds. The intensity of the storm is simulated as a Markov process using an environment-dependent hurricane intensity model (Jing & Lin, 2019). In Jing and Lin (2020), the authors generated more than 55,000 TCs from 100 independent realizations of the genesis of TCs during the period 1979–2014. We randomly select a total of 3,690 TCs from 10 independent 36-year simulations of PepC, which intensify and reach TC strength, to determine the time and location of potential synthetic TCs for initializing TC-GEN.

3. TC-GEN: Data-Driven Synthetic TC Downscaling Approach

TC-GEN, our data-driven synthetic TC downscaling approach based on high-resolution ML-GWM, includes the following main steps.

1. Obtain the annual frequency as well as the date and locations of synthetic TCs from an environment-based TC genesis model;
2. Generate synthetic TC seeds using data-driven approach;
3. Integrate synthetic TC seeds into the background environment represented by global reanalysis data;
4. Run Pangu-Weather to simulate the complete TC life cycle, detecting and tracking the TC from simulation outputs;
5. Adjust TC intensity (maximum sustained winds) by an intensity bias correction model;
6. Retrieve the size of synthetic TCs (outer extent where the wind decreases to a specific threshold) from the simulated TCs.

The idea is shown in Figure 1 and we explain each step in detail in the following sections.

3.1. Annual Counts, Date and Locations of TC Seeds

The first step of TC-GEN is to determine the annual frequency and the date and location of the synthetic TCs. We use an existing statistical approach—the environment-based hierarchical Poisson genesis model, which is the genesis component of PepC (Jing & Lin, 2020)—to determine the annual number of synthetic storms, as well as when and where they originate over the ocean basin. This method has been shown to generate TC formations that align closely with observations, including interannual variations. When initializing Pangu-Weather, we use this approach to estimate number of synthetic TCs expected each month. We then randomly assign a specific date and time for each storm within the month of genesis.

3.2. Structured TC Seeds Generation Via Data-Driven Approach

Previous studies using either statistical-deterministic or purely statistical TC downscaling approaches provide only the timing and position of TC seeds. There is no assumption about the structures or spatial features of the synthetic TCs at genesis. To create synthetic TC seeds with fine-grained spatial features, we employ a two-step process. First, we use PCA to learn a linear representation of atmospheric and surface variables from historical TCs at their genesis. PCA helps identify the most significant patterns in the data. Next, we create synthetic TC seeds based on this representation by sampling various sets of linear blending weights, according to the variance in the historical data.

PCA is a widely recognized method used for dimension reduction in the fields of data science and machine learning. Given high-dimensional data represented by a matrix X with dimensions $M \times N$, where M is the number of observations and N is the dimension of features, the main objective of PCA is to identify a set of orthogonal axes (principal components) along which the variance of the data is maximized. By projecting the original data onto these principal components, PCA effectively reduces the complexity of the data, while retaining the essential information contained in the data set. In a low-rank system, a small number of principal components are sufficient to explain the majority of the variance present in the data. Consequently, in such systems, it is feasible to generate synthetic data with only a few key principal components. PCA has been successfully applied in many earth science applications (Bretherton et al., 1992; Nandi et al., 2016), and we refer the reader to Abdi and Williams (2010) for more details.

We use all 690 historical TCs documented in IBTrACS during the period 1979 to 2022 to generate a linear representation of TC seeds using PCA. A TC seed is defined as a circular region with a radius equivalent to 64 grid cells at genesis (approximately 1,600 km), centered on the storm's eye. This circular region is large enough to encompass the outermost extent of most TCs at genesis. To train the model, we collect TC seeds for each of the 69 atmospheric and surface variables from the ERA5 reanalysis for all $M = 690$ storms. For each variable, we extract a square region that measures 128×128 grid points with the storm location as the center. This extraction process results in a matrix with dimensions of $128 \times 128 \times 69 \times 690$, representing the historical TC seed regions. Next, we perform PCA to reduce the dimensions of the data. The cumulative sum of the 20 largest eigenvalues is shown in Figure 2. We show that historical TC seeds form a low-rank system, with the top 10, 15, 20 principal components sufficient to explain more than 94.1%, 96.2%, and 97.2% of the variance in the data. Moreover, the top 50, 100 and 500 principal components explain more than 99% of the variance, with the top 500 principal components explaining more than 99.9%. Using 50, 100 and 500 principal components, we show that the reconstructed wind fields (including total wind, as well as the u and v components) effectively preserve most of the detailed spatial features present in historical TC wind fields, as in Figure 2. As the computation of TC reconstruction increases only linearly with the number of bases, and to include fine-scale local details, here we use 500 principal components that represent the majority of the variations in actual environmental fields. In doing this, we reduce the dimensionality of the raw data from $128 \times 128 \times 69$ to 500, creating a low-dimensional space for TC seeds.

Based on the top 500 principal components, we create synthetic TC seeds (the first step of the TC life cycle) by randomly sampling the low-dimensional space. We only sample one realization each time to represent one synthetic TC. The weight of each principal component is determined by randomly sampling from a Gaussian distribution centered around zero, with a variance equal to its corresponding eigenvalue (i.e., the variance explained). This allows synthetic TC seeds to have the same distribution as observational TC seeds, which is important for ensuring that synthetic TCs exhibit similar statistical characteristics to historical storms later in their life cycle. In Figure 3 we show three synthetic TCs at their genesis. We show four important variables (mean sea

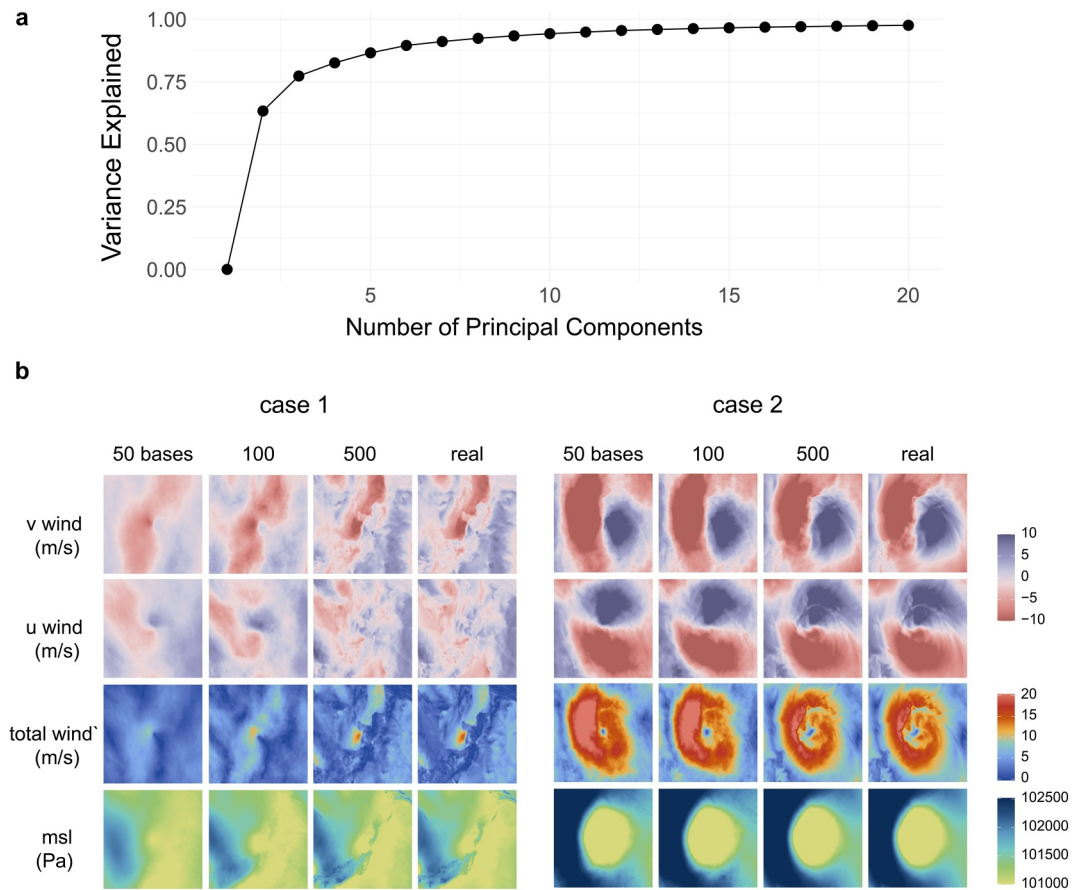


Figure 2. Reconstruction of tropical cyclone (TC) genesis using principal component analysis (PCA). We run PCA on TC historical data to learn a low dimensional representation for TC seeds at genesis. Subplot (a) shows the cumulative sum of the eigenvalues sorted in descending order, where each point represents the proportion of total variance explained by considering the corresponding number of top principal components. Subplot (b) shows the reconstruction of TC seeds at genesis for two distinct historical storms (referred to as case 1 and case 2), using 50, 100, 500 principal components, respectively. Both TC seeds are almost completely reconstructed using 500 principal components.

level pressure, temperature, and surface u- and v-winds) to demonstrate the effectiveness of this approach in simulating realistic TC seeds at genesis with diverse characteristics in terms of intensity, radius, and asymmetry.

3.3. Integrating Synthetic TC Seeds Into Background Environment With Weighted Poisson Blending

After generating a synthetic TC seed, we integrate it into the surrounding environment at a specific date and location identified by the PepC genesis component (see Section 3.1). This step is critical for Pangu-Weather in simulating environment along with synthetic TC seeds. To achieve seamless integration of the TC seeds with the background environment, we employ Poisson blending, a widely used approach to smoothly insert one image into another, without introducing artifacts at the boundary of the inserted image (Pérez et al., 2003). The fundamental concept of Poisson blending is to copy the gradients between spatially neighboring pixels rather than to directly copy the absolute color values. This approach ensures a smooth transition between different regions of the images, effectively eliminating visible seams and enhancing the coherence of the blended result.

Poisson blending can be mathematically expressed as follows: Let f^* be a known function defined on a closed subset of $S \subset R^2$ (e.g., color defined on pixel grids), and let g be another known function defined on a closed subset $\Omega \subset S$ with boundary $\partial\Omega$. The objective of poisson blending is to find an unknown function f on Ω , such that

$$\min_f \iint_{\Omega} |\nabla f - \nabla g|^2, \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega} \quad (1)$$

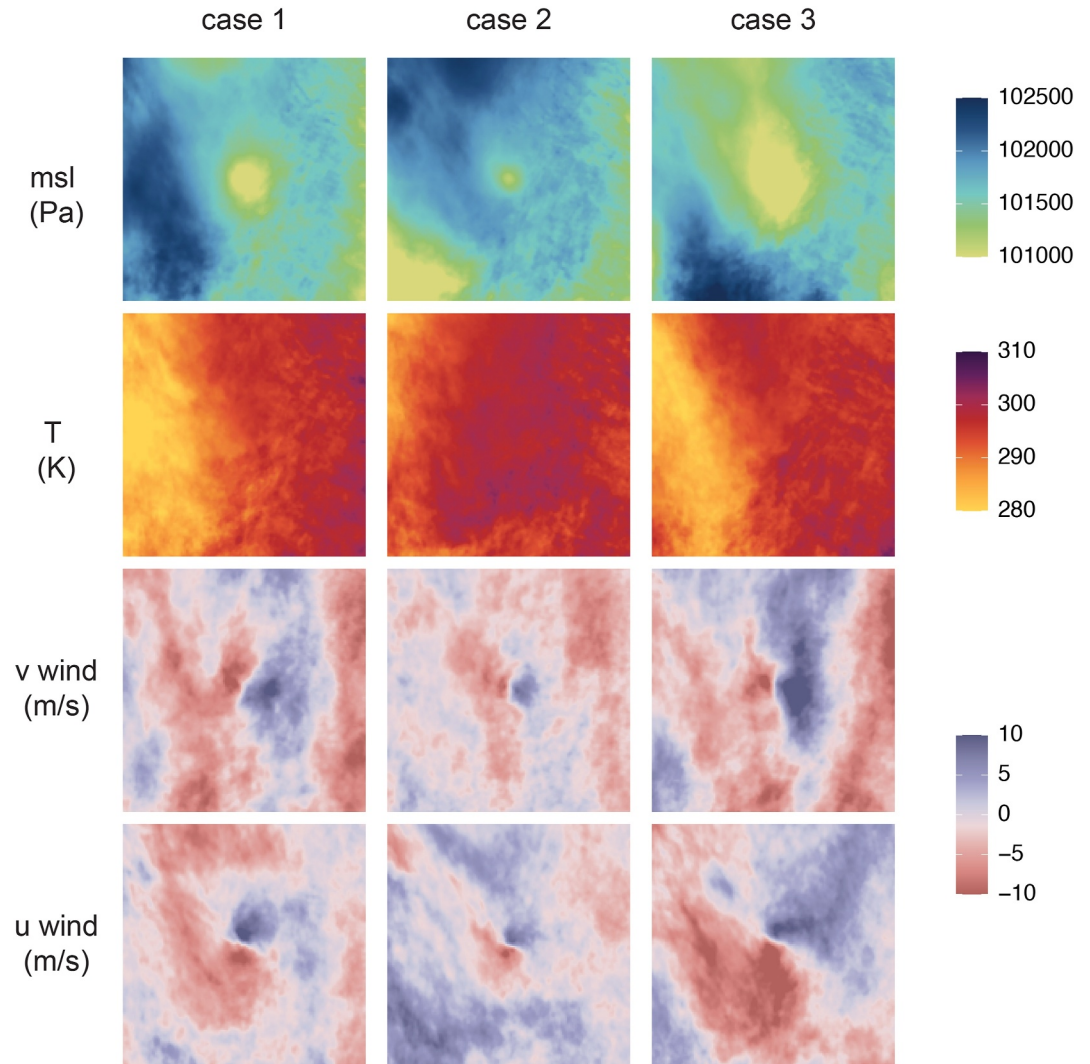


Figure 3. Create synthetic tropical cyclone (TC) seeds with 500 principal components. This figure shows three synthetic TC seeds at genesis using the data-driven approach with 500 principal components. Each column shows one case, and the four rows represent mean sea level pressure, temperature, surface v- and u-winds.

where $\nabla = \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right]$. This optimization aims to within the source region, it closely resembles the gradient of g , while at the boundary it should be similar to f^* . For more details and numerical solutions, we refer the reader to Pérez et al. (2003).

As an analogy to our problem, f^* and g represent the global environment map (i.e., target matrix) and synthetic TC seeds (i.e., source matrix), respectively. In this context, Poisson blending aims to combine these two matrices, ensuring that the gradients of the blended results within the source region closely resemble the synthetic TC seeds, while at the boundary, they should be similar to the background environment. In practice, we find that naive Poisson blending still results in noticeable artifacts. Therefore, we adopt a linear blending technique that incorporates both the source and target gradients, with a weight determined by the distance to the boundary. The optimization is modified to accommodate this approach:

$$\nabla = (1 - \lambda)\nabla f^* + \lambda\nabla g, \lambda = \begin{cases} \frac{5d}{R}, & d < \frac{R}{5} \\ 1, & \text{else} \end{cases} \quad (2)$$

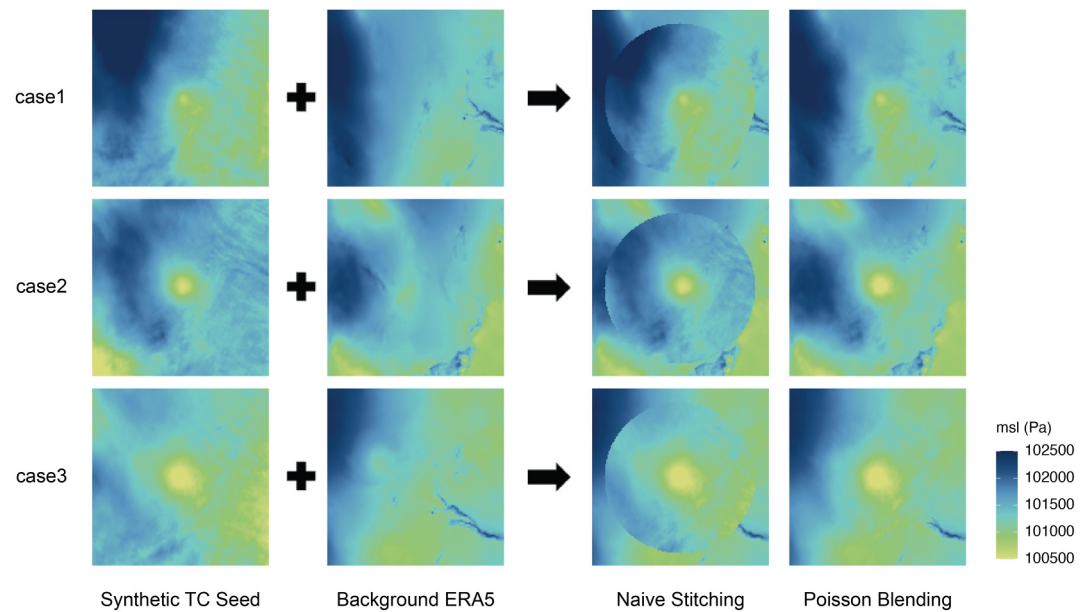


Figure 4. Tropical cyclone (TC) seed integration using Poisson blending. This figure shows the advantage of Poisson blending over naive stitching when integrating synthetic TC seeds into the environment. Mean sea level pressure is used as an example due to its ability to clearly reveal the “low-pressure center” of the storm. Naively stitching results in a sharp and unrealistic boundary. In contrast, the Poisson blending approach effectively integrates TC seeds into the environment, preserving both structures and achieving a smooth blend.

where d is the distance of each pixel to the boundary and R represents the radius of the blended region. As shown in the equation, within the outer 1/5 of the circular blending region, the gradients transition from the environment to the seed, while in the inner 4/5 of the domain, the seed gradients are used. This optimization ensures a smooth integration of TC seeds into the global environment map, maintains natural transitions at the boundaries while preserving the structure of the TC seeds, without sacrificing detailed features. We note that the radius of the blended region should be carefully chosen so that the statistics of the TC seeds (such as the intensity of the TC seed) remain consistent before and after the blending (Figure S1 in Supporting Information S1).

We use this Poisson blending approach to seamlessly integrate the 69 atmospheric and surface variables of synthetic TC seeds within a radius of 64 grid cells (approximately 1,600 km) into the corresponding ERA5 reanalysis data. The intensity of synthetic TC seeds is consistent before and after blending (Figure S1 in Supporting Information S1). In Figure 4, we show the Poisson blend process using mean sea level pressure as an example, which clearly reveals the low-pressure center of the storm. Similarly, we show three cases as examples of weak, moderate, and strong storms, illustrating the effectiveness of Poisson blending across a range of storm characteristics.

3.4. Simulation and Tracking

After seamlessly integrating the generated TC seeds into the background environment map, we proceed to run Pangu-Weather, simulating the entire life cycle of each storm over a 15-day period, with 6-hr intervals between each step. To track the location of storm center from Pangu-Weather outputs, we formulate a Gaussian kernel over mean sea level pressure and a Laplacian kernel over vorticity at 10 m, to identify the local maximum vorticity or minimum pressure associated with the characteristic bell-shaped symmetric structure typical of a storm. Then we determine the center of the storm by averaging the locations of these two local extrema. The Laplacian kernel is derived from the subtraction of two Gaussian kernels with different variances. This specific kernel shape is particularly responsive to circular-like stimuli, such as the vorticity structure of TCs. Empirically, we set $\sigma = 2$ for the Gaussian kernel and $\sigma_1 = 2$, $\sigma_2 = 8$ for the Laplacian kernel, for the best performance. In most cases, the positions of maximum vorticity closely align with those of minimum sea level pressure. However, in particular situations, such as when storms approach small islands, this approach offers more reliable tracking compared to solely relying on the pressure field (such as the TempestExtremes method, which is widely utilized for identifying

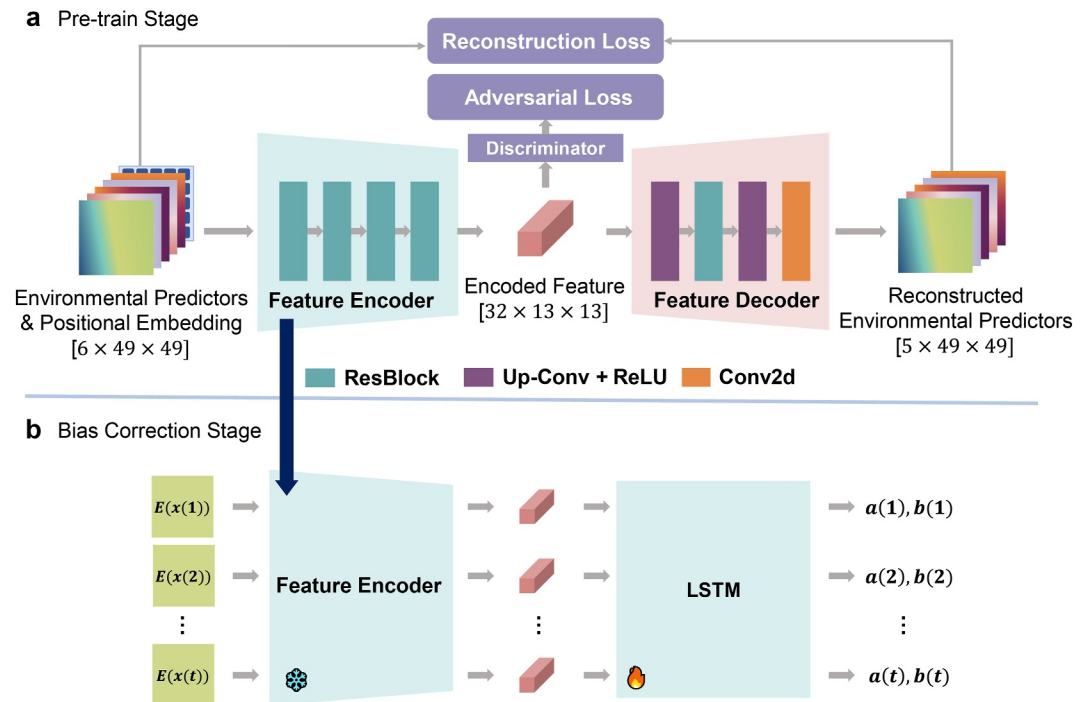


Figure 5. Structure of the intensity bias correction model. The model consists of two stages, (a) a pre-trained stage to bridge domain gap between ECMWF Reanalysis v5 and Pangu-Weather output so that all features become domain-agnostic; and (b) a bias correction stage that adjust the maximum wind speed using a long short-term memory model.

storm locations in gridded climate data (Ullrich & Zarzycki, 2017; Zarzycki et al., 2021). Also, Pangu-Weather includes an algorithm for tracking TCs, which is based on relative vorticity, geopotential thickness, and 10-m wind speed. Our tracking method delivers comparable results, while requiring fewer data inputs. This advantage makes our approach generalizable to other weather forecasting systems that may lack specific variables (e.g., ForecastNet does not output vorticity; see Pathak et al., 2022).

3.5. Bias Correction of TC Maximum Wind Speed and Minimum Central Pressure

The Pangu-Weather model is trained using ERA5 reanalysis data. Since the fine-grained structure of a storm cannot be fully resolved in the ERA5 reanalysis data, the physical processes of simulated storms are also not correctly resolved, leading to an underestimation of the real intensity of the storm. To address this problem, we develop a separate machine learning model to correct for this bias based on the characteristics of the storm and the environment within the inner region of the storm. We correct for both maximum wind speed and minimum central pressure simultaneously, with the model outputting both variables, as minimum central pressure is considered a better predictor of economic damage (Klotzbach et al., 2022). Due to the recurrent nature of this problem, where the wind speed at time t is highly correlated with the wind speed at both the previous and future time steps, we formulate the problem by a recurrent neural network based on bidirectional long-short-term memory (LSTM), using the evolution of real intensity throughout the lifetime of TC to develop the model. The structure of our network is shown in Figure 5 subplot b, where the bias correction stage is formulated as.

$$\mathbf{I}_{\text{true}}(t) = e^{a(t)} \mathbf{I}_{\text{raw}}(t) + b(t) \quad (3)$$

$$a(t), b(t) = \mathbf{F}(\mathbf{E}(\mathbf{x}(1)), \mathbf{E}(\mathbf{x}(2)), \dots, \mathbf{E}(\mathbf{x}(t)) | \theta) \quad (4)$$

In Equation 3, we use the term “raw” to indicate the intensity directly simulated from TC-GEN, which is expected to have the same statistics as those in the ERA5 reanalysis data. The term “true” is used to denote the real intensity, which has statistics similar to IBTrACs. Thus, \mathbf{I}_{raw} and \mathbf{I}_{true} represent the maximum wind speed or minimum central pressure before and after bias correction. We use an exponential function to transform $a(t)$ to

ensure the scaling factor remains positive, which helps the model learn the relationship more easily and improves model performance. $\mathbf{E}(\mathbf{x}(t))$ represents the environmental predictors extracted from the area surrounding the TC center, which is located at $\mathbf{x}(t)$ at time t . \mathbf{F} represents a machine learning model with learnable weights θ .

We select five environmental variables (mean sea level pressure, u- and v-component of 10 m wind speed, relative humidity at 850 hpa and temperature at 850 hpa) that have been identified as the most important predictors of the intensity of TC (Jing & Lin, 2019). At each time step, we use the environmental fields of these five predictors within a circular region that covers a radius of 49 grid cells to create the raw input. Thus, the raw inputs have a size of $5 \times 49 \times 49$ (Lam et al., 2022). We then convert the raw input to a feature vector using a feature encoder comprised of 4 ResBlocks, and feed these feature vectors into an LSTM, predicting $a(t), b(t)$ at each time step.

In practice, we note that a model trained with ERA5 as input may not perform equally well during the test phase when the inputs are derived from Pangu-Weather predictions. This discrepancy arises because of the well-known domain gap issue in deep learning (Nam et al., 2021; Tremblay et al., 2018; Wei et al., 2018). Essentially, nuanced differences between training and testing data on the input side (Pangu-Weather simulation vs. ERA5 in our case) can lead to a catastrophic drop in model performance.

To bridge the domain gap, we pre-train the feature encoder so that the extracted feature is informative to reconstruct the raw ERA5 environment, yet indistinguishable in terms of its source, that is, whether it comes from ERA5 or Pangu-Weather. Such properties are achieved by training the feature encoder through an auto-encoder architecture, as shown in the Figure 5 subplot a. Specifically, compressed feature vectors go through a feature encoder, followed by a feature decoder to reconstruct the original features using an L1 loss. Additionally, we introduce an adversarial loss that performs a binary classification task, attempting to discern the source of the feature (ERA5 or Pangu-Weather). The feature encoder is trained to deceive the discriminator to the extent that it cannot identify the source of the feature. As a result, the extracted features become domain-agnostic after this stage and the domain gap is mitigated.

Overall, we start by training the auto-encoder architecture using environment maps randomly sampled from both the ERA5 and Pangu-Weather output (Figure 5 subplot a). The auto-encoder is designed to reconstruct its input, whether it comes from ERA5 or Pangu-Weather. By exposing the auto-encoder to data from both domains during training, it is optimized so that the latent features are indistinguishable between the two sources. Once the auto-encoder converges, we discard the decoder, freeze the encoder weights for feature extraction, and only update the LSTM weights (Figure 5 subplot b). To train the LSTM, we divide the historical data of the TCs from 1979 to 2021, randomly selecting 80% of the TCs (547 TCs) for training and reserving the remaining 20% (139 TCs) for testing. We use the real maximum wind speed and minimum central pressure data from IBTrACS as ground truth. The LSTM is trained with an AdamW optimizer (Loshchilov & Hutter, 2017), with a Huber loss and a learning rate of 0.0001 over 10 iterations. The model achieves a root mean square error (RMSE) of 8.1 m/s on training sets and 7.8 m/s on testing sets, which are fairly close and indicate no overfitting during training.

The performance of the LSTM-based bias correction model is shown in Figure 6, which shows four cases that represent various characteristics of the storm. These include typical growth and decay, rapid intensification, and storms that weaken after hitting an island but subsequently strengthen after moving over the ocean. More cases, as well as the bias correction for minimum central pressure, are shown in Figure S2 in Supporting Information S1.

3.6. Extracting the Radius of Outer Size From Each Storm

The destructive potential of a TC is related to both its maximum sustained wind speed and the radial extent of the near-surface wind, the latter typically measured by the outer size of the TC (Irish et al., 2008; Powell & Reinhold, 2007). Therefore, it is important to assess how well TC-GEN can replicate the realistic horizontal wind structures of TCs. Furthermore, since TC seeds are artificially integrated into the surrounding environment, successfully reproducing the historical distribution of the outer size of TCs can, to some extent, prove that integration based on weighted Poisson blending is feasible.

Previous studies have shown that reanalysis data sets, including the ERA5 reanalysis used in this study, are capable of effectively capturing the TC outer size, which is defined by radii where convection is minimal and the atmosphere maintains a radiative–subsidence balance (Schenkel et al., 2017). In earlier studies, multiple thresholds for azimuthal mean tangential winds have been used, such as 2, 6, 8, 9, and 12 m/s (Bian et al., 2021).

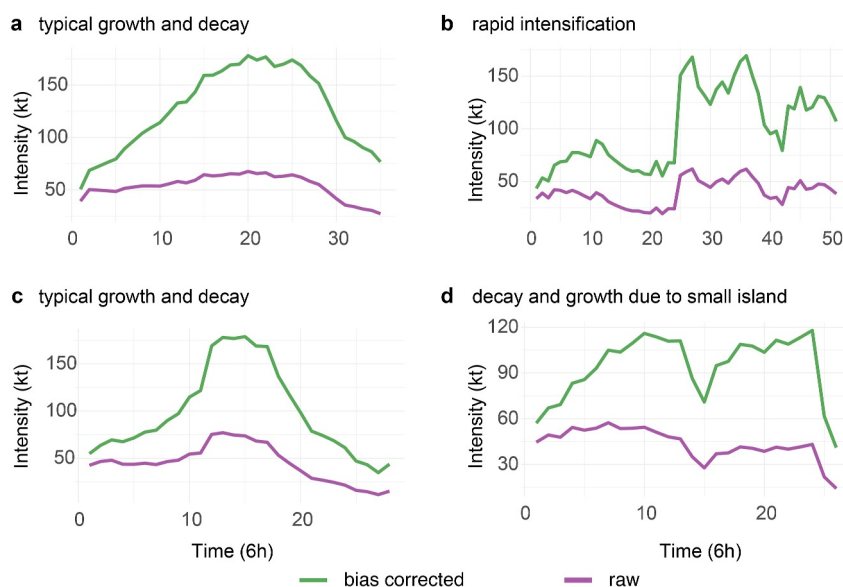


Figure 6. Examples of bias correction in tropical cyclone (TC) maximum wind speeds. Four synthetic TCs are shown including (a, c) two TCs illustrating a storm's typical growth and decay, (b) TC underwent rapid intensification, and (d) TC hit small islands followed by subsequent growth. The bias correction model effectively simulates the realistic evolution of intensity while capturing the correlation of TC intensity with the previous time step, thus maintaining strong continuity.

Here, we use three size metrics based on the radii where the wind diminishes to 2 m/s (r_2), 6 m/s (r_6), and 12 m/s (r_{12}), to establish a range of size metrics that include both the smallest and the largest metrics.

We identify the outer size of the TC following the procedures outlined in (Schenkel et al., 2017). The procedure involves the following five main steps: (a) remove the background flow from the surface wind field, which is empirically defined as the storm translation vector rotated 20 deg cyclonically and diminished by a factor of 0.55, as described in the method by Lin et al. (2012); (b) establish a TC-centered polar coordinate system, where the cyclone center serves as the origin, and the grid spacing in the radial direction is set at 0.5 times the archived reanalysis grid spacing; (c) compute the tangential wind at the grid points of the reanalysis data for each annular region of the polar coordinate system; (d) remove all radial points if the data asymmetry parameter exceeds 0.5, following Chavas and Vigh (2014); (e) calculate the gridded radial profile of azimuthal-mean 10-m tangential wind using cubic Hermite polynomial interpolation, based on which we then identify the radii of at which the 10-m azimuthal-mean tangential wind speeds decrease to a certain threshold (12, 6, and 2 m/s).

4. TC-GEN Evaluation

We evaluate the performance of TC-GEN by comparing simulated TCs with historical observations and two existing TC downscaling approaches (KE08 and PepC) across various TC characteristics, including track density, landfall frequency along the US-Mexico coastline, LMI, landfall intensity, and outer size under three size metrics. For each TC, we generate a synthetic TC seed and blend it into the corresponding hourly or monthly environment map, with the location and time provided by the PepC genesis model. We then run Pangu-Weather to simulate each storm and apply the bias correction model to adjust the maximum intensity for each step. The track is terminated if the raw maximum intensity is below 8 m/s, the adjusted intensity is below 15 kt, the central vorticity is below $5e^{-5} s^{-1}$, or if the storm has been over the land for five consecutive steps, which is equivalent to 30 hr. To form a fair comparison, we remove storms with a LMI less than 25 kt for all data sets. As TCs would undergo an extratropical transition at high latitudes, we restrict our analysis to samples where the storm center is south of 50 N. The remaining storms are used for the TC-GEN evaluation.

4.1. Track Density

Figure 7 compares the simulated tracks that are initiated with both hourly and monthly data, with observed tracks and simulated tracks using KE08 and PepC. The colors represent the spatial track density normalized to the

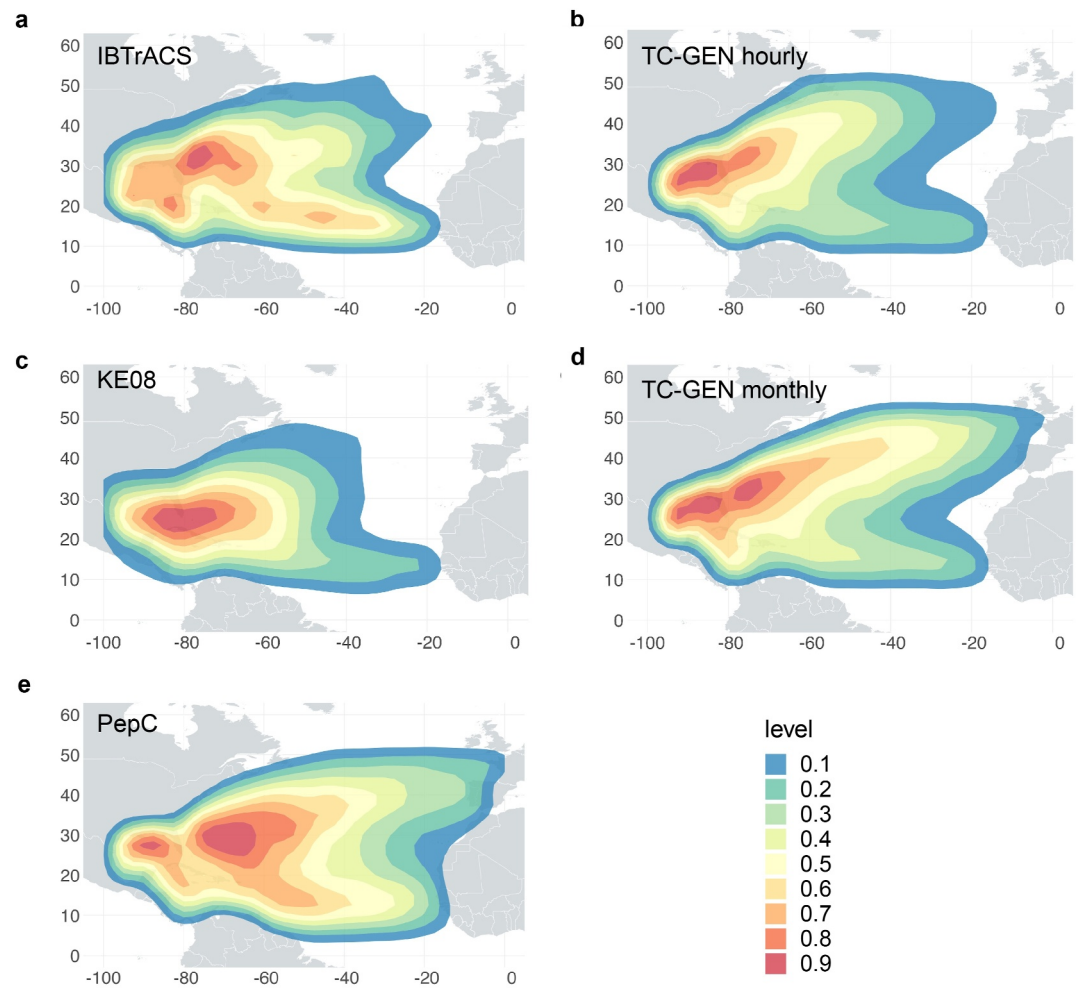


Figure 7. Track density Track density is calculated as the accumulated number of tropical cyclone passes into each 0.75×0.75 grid box normalized by the maximum grid value of the basin, smoothed with a Gaussian low-pass filter. Tracks over land are removed in each subplot. Simulated tracks initiated with (b) hourly data and (d) monthly data are compared with (a) observations, (c) KE08 and (e) PepC. All simulated tracks replicate the typical recurving pattern seen relatively well.

maximum of the basin. We show that both sets of simulated tracks replicate the typical recurving pattern seen in the observed tracks relatively well, which is comparable to that of KE08 and PepC. Compared to observations, TC-GEN simulated tracks exhibit a negative bias in the main development region (Figure S3 in Supporting Information S1). Considering that the genesis process of TC-GEN heavily relies on the genesis component of PepC, while the track and intensity components are deterministic, this bias could mainly stem from the negative bias in PepC within this region. We test this hypothesis with a sensitivity test by resampling the genesis according to the spatial distribution of historical genesis locations. After sampling, we find that the simulated tracks effectively capture the hotspots in the main development region, the southeast US coast, and the Gulf of Mexico, although there is a slight positive bias in the Gulf of Mexico with hourly data and in the West Caribbean Sea with monthly data (Figure S4 in Supporting Information S1).

We further evaluate the performance of TC-GEN by comparing the 6-hourly north-south and east-west displacements of the simulated tracks with those simulated by KE08 and PepC, as well as observed tracks in IBTrACS. This comparison serves as a means to assess the performance of TC-GEN in simulating individual tracks. The results are shown in Figure 8. All simulated data sets are largely consistent with the observations. For simulated tracks initiated with hourly data, there is a slight positive bias for positive meridional displacement, a slight negative bias for negative meridional displacement; and correspondingly, a negative bias for positive zonal displacements and a positive bias for negative zonal displacements, which could come from fewer recurving

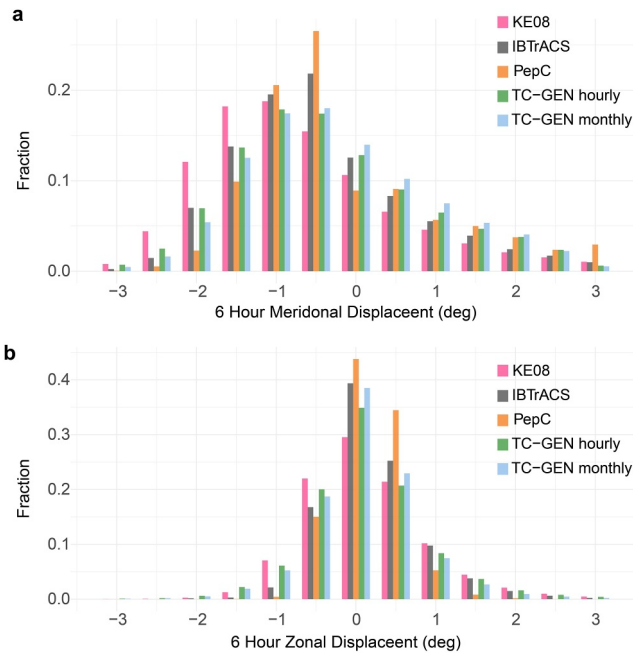


Figure 8. 6 hour zonal and meridional displacement Comparison of probability density functions of 6 hr (a) meridional and (b) zonal displacements between simulated tracks and observations. Simulated tracks initiated with hourly data and monthly data are both presented and compared with KE08, PepC and IBTrACS.

tracks in simulated storm tracks that originate in the main development region. Additionally, positive biases in meridional displacements may also arise from the deviations in eastward moving tracks at high latitude (near Europe), where they should have been terminated because of their low wind intensity. These patterns are also seen in simulated tracks that are initiated with monthly reanalysis data. In general, both data sets exhibit comparable performance to existing downscaling methods, with simulated tracks initiated using hourly data showing slightly better performance than those initiated with monthly data.

4.2. Landfall Frequency

We examine the annual regional landfall frequency at coastal locations along the North Atlantic coastline. To help indicate locations, a total of 186 mileposts (MPs) are defined following Vickery et al. (2000), as shown in Figure 9, to cover the coastline with 100-km spacing along the Mexican coastline and 50-km spacing along the US coastline. Landfall is defined as simulated or observed storms that approach within 50 km of each coastal milepost. The results presented in Figure 9b are based on simulated tracks initiated using both hourly and monthly data. Furthermore, results from KE08 and PepC-simulated tracks are included for comparison. As a reference, the historical annual landfall rate between 1900 and 2022 is shown with shading that indicates the associated sampling error at each milepost. The sampling error for the annual rate of each gate is determined by first calculating the standard error of the number of storms that cross each gate over the entire period, and then dividing by the number of years.

Due to the different annual total frequencies in the simulated and observed track data sets, which can significantly influence landfall frequency, we adjust the annual rate to a uniform 13 storms per year in all data sets to form a fair comparison. Our results indicate that the simulated tracks, from daily and monthly data, can reproduce the overall pattern in the observations, showing a performance comparable to that of KE08 and PepC. The annual landfall rates simulated by TC-GEN exhibit correlations with observations of 0.79 and 0.82 for the tracks initiated with hourly and monthly data, respectively, which are comparable to 0.83 and 0.79 for PepC and KE08, respectively. Both TC-GEN simulated data sets capture the landfall frequency at MP 100–125, which is overlooked by existing methods, while they underestimate the landfall frequency below MP 25, possibly due to fewer genesis in the main development region.

4.3. Lifetime Maximum Intensity and Landfall Intensity

We analyze the simulated TC intensity using two metrics, the LMI and landfall intensity. The LMI distribution serves as a representation of the TC intensity climatology. A successful simulation should be able to reproduce the bimodal distribution with a shoulder feature around 120 kt, which is associated with storms that undergo rapid intensification (C.-Y. Lee et al., 2016). We show in Figure 10 that after bias correction, both sets of TC-GEN are capable of reproducing a realistic distribution of the observed LMI. TC-GEN tracks initiated with hourly data better simulate the tail of the LMI distribution for LMI greater than 75 kt; however, there is a negative bias in LMI for storms with LMI less than 75 kt, which are mostly moderate storms that do not undergo rapid intensification. TC-GEN tracks initiated with monthly data overestimate the proportion of storms with an LMI greater than 75 kt. The biases might originate from the intensity bias correction model. In order to have a reasonable fraction of storms undergo rapid intensification, the bias correction model prioritizes strong storms, which could lead to a higher proportion of storms becoming more intense than they should be. Similar patterns are also seen in the distribution of the landfall intensity. Both data sets initiated with hourly and monthly data exhibit a positive bias in landfall intensity greater than 75 knots. This bias is also likely to be attributed to the intensity correction model, which tends to overestimate the intensity of the storm when the storm has already weakened.

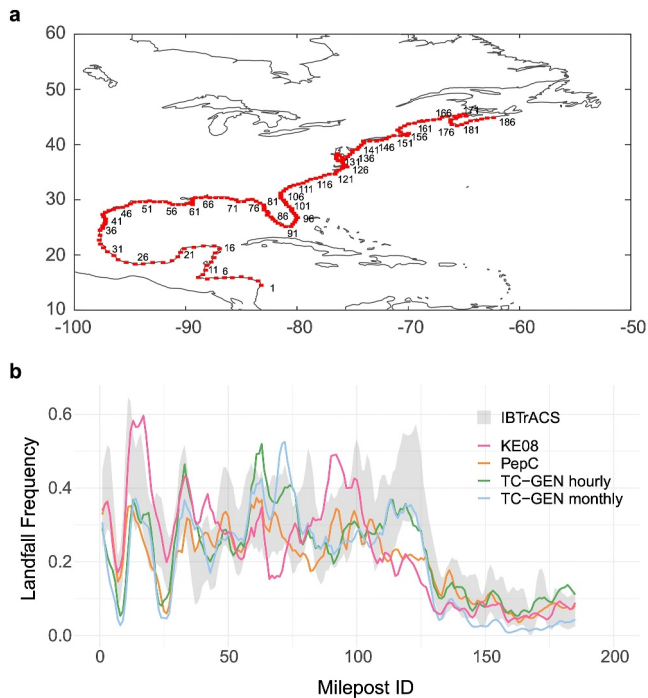


Figure 9. Landfall frequency Subplot a shows locations of mileposts along Mexico (every 100 km) and U.S. (every 50 km) coastline. Subplot b shows the comparison of annual landfall rate at each of 186 mileposts between simulated tracks and observation, with shading indicating the associated sampling error. Results from KE08 simulated tracks and PepC simulated tracks are shown for comparison.

4.4. Outer Size

Although the TC structure cannot be fully resolved in the reanalysis data set, previous work has shown that the reanalysis data can be used to extract the outer size of storms. Recent studies have documented that ERA5 exhibits an improved representation of the outer size compared to previous versions of ERA reanalysis (Bian et al., 2021), including a high correlation and low RMSE compared to observations derived from satellite remote sensing (Gori et al., 2023). Here, we analyze the comparison of the outer size of TCs between TC-GEN-simulated storms and historical storms. We randomly select 700 synthetic TCs whose seeds are generated using the PCA analysis, and simulate the entire life cycle of each storm using TC-GEN. We then extract the outermost size of the storm at each 6-hr time step and collect storm sizes from all steps (not only the first step) to compare with historical TC sizes that are derived from the corresponding ERA5 reanalysis data. In this comparison, we only focus on storms that occur over the ocean, and future work may further investigate how well storm sizes are represented after making landfall. As the statistical downscaling approach does not provide this output, we do not include data from KE08 and PepC for this analysis.

The results are shown in Figure 11. The medians (standard deviations) of r_2 , r_6 , and r_{12} from TC-GEN simulated tracks are 574.5 (230.2), 451.0 (172.5) and 270.8 (74.1) km, respectively, compared to 584.9 (259.8), 468.1 (233.4) and 277.5 (129.6) km, respectively, from historical data. For all three size metrics, TC-GEN accurately reproduces the median of the outer size, with a discrepancy of approximately 10 km, which is even lower than the uncertainties in the outer size arising from different reanalysis data sets (Bian et al., 2021). However, the standard deviations for all three size metrics in TC-GEN-simulated TCs are smaller than those observed in historical storms, particularly for r_{12} , which represents the largest metric that defines the outer

size of storms. This smaller variation may arise from the bias in the horizontal wind structures of TCs at genesis, which requires further investigation. Note that the consistent size distribution does not necessarily indicate that TC-GEN can perfectly model TC size, but a strong indicator that our TC seed generation and sampling algorithms are highly bias-free.

5. Discussion

5.1. Sensitivity to Reanalysis Data

In this study, we show the results of TC-GEN initiated with 6-hourly and monthly ERA5 reanalysis data. We illustrate that TC-GEN can generate realistic synthetic TCs with both data sets, despite their different temporal resolutions. Given the significant advantage in reducing data download and storage burdens, we propose that using monthly data to initiate TC-GEN is acceptable when scaling up the data generation process to achieve global coverage or long-period simulations.

Furthermore, we evaluate the suitability of applying TC-GEN to reanalysis data sets other than ERA5 reanalysis. Using NCEP GFS (see Section 2.3, we find that the results are unsatisfactory (as illustrated in Figure S5 in Supporting Information S1), probably attributed to the well-known domain gap problem in deep learning (Nam et al., 2021; Tremblay et al., 2018; Wei et al., 2018). In a deep learning model, typically the first several layers of the neural network are responsible for extracting features from raw data. These layers are trained to identify relevant features that help in downstream tasks, such as recognizing distinctive patterns, reducing noise, and achieving specific invariances (Zeiler & Fergus, 2014). However, these layers can become highly specialized for the specific training data set and fail to generalize effectively to different data sets. We believe that this is the reason why TC-GEN does not perform well in NCEP GFS, as its core model Pangu-Weather is trained exclusively on ERA5 reanalysis data. Therefore, when using TC-GEN for synthetic TC downscaling under future climate projections, it is also necessary to pretrain and redevelop the core simulator (ML-WMG) to eliminate any domain gaps. Building upon this, another interesting future direction also involves the design of a generalized

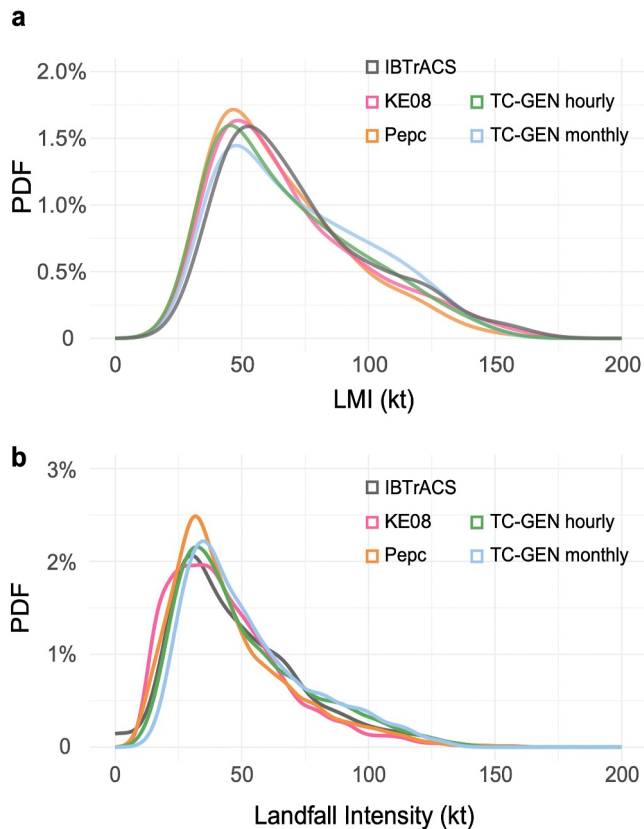


Figure 10. Lifetime maximum intensity (LMI) and landfall intensity Subplot a compares LMI distribution between simulated storms and observations. Subplot b compares the distribution of landfall intensity, defined as the maximum wind speed within 50 km of each coastal milepost, from both simulated and observed tracks. Results of KE08 and PepC are shown here for comparison.

ML-based weather forecasting system that performs reasonably well on various reanalysis data sets. This could be accomplished, for example, by training the model on multiple reanalysis data sets together.

5.2. Extrapolation and Applicability for Future Climate

To generate storms under future climate conditions, traditional statistical downscaling approaches assume that the observed relationships between the TC and the environment, established under historical climate conditions, will continue to hold in a warming climate (Emanuel et al., 2008; Jing et al., 2021). While there is no need to re-train the statistical models based on future climate, the process of extrapolation introduces a certain degree of uncertainty, particularly when unforeseen factors may impact the relationship. TC-GEN has unique strengths and limitations in addressing extrapolation. As ML-GWM operates in a manner that mimics the characteristics of numerical models; once a future ML-GWM becomes accessible, it enables TC-GEN to directly simulate storms under future climate conditions when initiated with future climate projections. However, as previously discussed and a significant limitation, the optimal performance of TC-GEN depends on being paired with the specific set of environment maps on which it was trained. Therefore, it is necessary to pre-train an ML-GWM using environmental data obtained from climate projections, before applying TC-GEN to climate change studies.

5.3. Range of TC Seeds

Poisson blending involves integrating synthetic TC seeds into the surrounding environment, where it is crucial to carefully choose a specific range to define the extent of TC seeds. The TC seed range should be limited to avoid including unrelated meteorological systems, which can negatively affect the performance of the PCA model. However, if the range is excessively limited, it may not fully capture the entire spatial structure of the TC during its genesis. In such cases, the outermost extent of the range may not have fully diminished to the intensity of the surrounding background wind. As the Poisson blending algorithm tends to assimilate the gradient of TC seeds while

aligning the boundaries with the ambient wind field, this could result in an underestimation of the wind speed within the inner region of the storm. In practice, we examine several sets of radii ranging from 25 grid cells (approximately 350 km) to 89 grid cells (approximately 1,200 km). Our sensitivity tests reveal that the results are relatively robust when the radius falls within the range of 49–75. For the primary analysis in this study, we use a radius of 64 grid cells as the optimal parameter for blending.

5.4. Intensity Bias Correction

The spatial structure of the TCs could not be fully resolved on the $0.25^\circ \times 0.25^\circ$ horizontal grid, and we apply the intensity bias correction model to adjust the simulated raw intensity to the real intensity. The bias correction model has the ability to reproduce a realistic LMI distribution and captures the correlation of the TC intensity with the previous time step, thereby maintaining strong continuity.

The machine learning-based intensity bias correction model has a more complex structure compared to the intensity model of existing statistical TC downscaling approaches. This is because the model not only corrects the underestimated intensity, but also bridges the domain gap between Pangu-Weather output and ERA5. In the development of this model, we tested models that are not based on recurrent neural networks, where the maximum intensity of the storm is adjusted solely based on the TC and environmental predictors in the current and previous steps. However, the performance of these models is not optimal, indicating that recurrent networks are necessary. We also experimented with models lacking a feature encoder and simpler regression methods (such as the intensity model in PepC, Jing & Lin, 2020), but they also fell short in bridging the domain gap. It is worth noting that

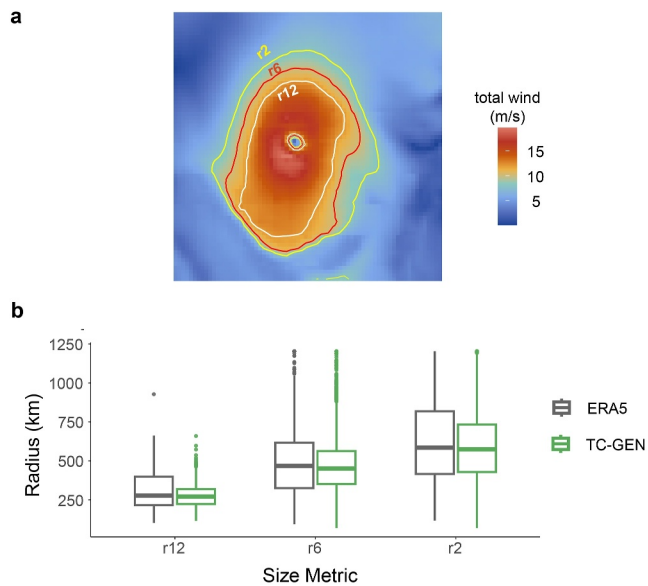


Figure 11. Tropical cyclone (TC) outer size Subplot a shows the concept of TC outer size defined as the radii at which the 10-m azimuthal-mean wind speed equals 2, 6, and 12 m/s. The contours of the 10-m azimuthal-mean wind are shown with color lines. Subplot b shows the boxplots of outer size at three radii metrics, from TC-GEN simulated tracks and those in ECMWF reanalysis v5. The median of each metric is represented as a horizontal bold line, and the upper and lower boundaries of each box indicate the 75th and 25th percentiles.

we have not ruled out the possibility that a simpler model could be equally effective, and future work may explore the development of such a model.

The interpretation of the intensity bias correction model should consider its limitations. One limitation is that it tends to overestimate the maximum wind speed of TC when the storm is decaying, which partially explains the positive bias in the landfall intensity distribution. This bias is not observed in the correction for the minimum central pressure. After bias correction, the central pressure of the storm increases to a normal level. The positive bias in maximum wind speed is likely a result of the model being trained to prioritize replication of the shoulder feature in LMI, which is associated with strong storms that undergo rapid intensification (this may also stem from a limitation in the Dvorak technique, which is used to estimate the TC intensity from satellite imagery; see Velden et al., 2006 for more details). Consequently, the model tends to generate more storms that grow at a higher rate while decaying at a slower rate.

Secondly, while the combination of Pangu-Weather and the bias correction model effectively simulate the overall intensity evolution of individual TC events, bias exists in the distribution of 24-hourly intensification rates (i.e., the change in maximum wind speed within 24 hr). As shown in Figure S6 in Supporting Information S1, TC-GEN, particularly when initialized with monthly ERA5 data, produces more instances of small and moderate intensification (or deintensification) compared to historical storms, while underestimating the fraction of rapid intensification (or deintensification) when compared to KE08 and PepC. This discrepancy arises because the bias correction model may overestimate storm intensity at various timestamps, reducing its effectiveness in simulating the transition from weak to strong

storms and leading to an underestimation of rapid intensification. Future work should focus on improving the bias correction model to effectively handle storms that undergo rapid intensification and those that do not, to achieve optimal performance for both LMI and landfall intensity.

Another limitation of the bias correction model is related to the well-known domain gap issue in deep learning (Nam et al., 2021; Tremblay et al., 2018; Wei et al., 2018). Specifically, if the climate data set is replaced with other reanalysis data or is derived from future climate model projections, the intensity bias correction model would require redevelopment. Although the current model structure, based on recurrent neural networks that utilize long-short-term memory, is effective, the model parameters would need to be re-trained to accommodate different climate data sets.

5.5. Asymmetric TC Wind Field

Horizontal asymmetric TC wind fields are important for disaster management and regional risk assessment (Wang et al., 2022). In addition to maximum wind speed, accurate estimation of the TC wind field is essential to identify at-risk populations and assess potential climate-related threats (Jing et al., 2024). One notable advantage of TC-GEN is its ability to directly output asymmetric TC surface winds. Although a comprehensive investigation of this capability exceeds the scope of this study, our basic evaluation of the asymmetries in the wind field related to TC motion indicates that TC-GEN demonstrates some preliminary capability in this regard (Figure S7 in Supporting Information S1).

Despite TC-GEN showing skill in simulating wind asymmetries, we note that the fine-scaled surface wind of simulated TCs may not be realistic. Nevertheless, utilizing the outer size of simulated TCs, which are well captured and well simulated in the reanalysis data, there are still various ways to effectively utilize the simulated asymmetries and leverage the asymmetric outer size. For example, with the adjusted maximum wind intensity after bias correction and the extracted outer size of the TC, a parametric wind model, such as the one developed by Chavas et al. (2015), can be applied to generate a complete asymmetric and continuous wind profile for the TC. Future work should also assess the ability of TC-GEN to simulate asymmetric winds at landfall. This assessment

may involve comparing the synthetic landfalling TCs with observational data, such as winds obtained from Automated Surface/Weather Observing Systems, or with output from dynamical simulations.

6. Conclusions

This study introduces “TC-GEN,” a novel machine learning-based approach for synthetic downscaling of TC, leveraging machine learning-based global weather model (ML-GWM). ML-GWM takes environmental maps of key atmospheric and surface variables from reanalysis data as input and predicts the values of these variables for the next time step, with the flexibility to choose from various lead times. Using a pre-trained model representing a cutting-edge ML-GWM (Pangu-Weather), we show that TC-GEN is capable of simulating synthetic storms that allow for the two-way interactions between the storm and its environment. It maintains computational efficiency that is similar to existing statistical-deterministic and statistical downscaling approaches; however, its distinctive advantage lies in the ability to simulate the spatial asymmetries of surface wind. TC-GEN consists of six key steps: the identification of the location and time of synthetic TC genesis, the generation of a synthetic TC seed for each storm through a data-driven process, the merging of the TC seed with the background environment using Poisson blending, the simulation of the full life cycle of the storm with ML-GWM, the correction of maximum wind speed and minimum central pressure biases using a long-short-term memory model, and the extraction of TC outer size from simulated wind fields. By comparing TC-GEN-simulated storms with observed storms and those simulated from existing statistical-deterministic and statistical downscaling approaches, we show that TC-GEN is capable of simulating storms that reproduce a range of important TC characteristics, including metrics for track, intensity, and storm size. For future work, our plan involves expanding this approach to include other TC basins and exploring its applicability on a global scale. Additionally, considering the absence of rain rate in the simulation output from Pangu-Weather, one potential work would be to examine the performance of TC-GEN when working with other ML-GWMs that are capable of simulating this important variable.

TC-GEN represents a fundamentally different TC downscaling approach compared to existing statistical-deterministic and pure statistical downscaling methods, and has the potential for further improvement with the advances in high-resolution ML-GWM that have higher resolution, improved accuracy, and even lower computational costs. Therefore, we expect that the performance of TC-GEN will see further improvements as ML-GWM undergoes continuous development. Moreover, using TC-GEN as an example of how recent advances in machine learning and data science can contribute to TC risk assessment, we believe that machine learning-based global weather models will play a crucial role in future climate studies.

Data Availability Statement

Pangu-Weather trained models can be downloaded from the public GitHub repository at (Bi et al., 2023b). Tropical cyclone observations are obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) project at (Knapp et al., 2018). Historical ECMWF reanalysis v5 data (both monthly and 6-hourly) are obtained from the ECMWF climate data store. Historical NCEP reanalysis data can be downloaded at (NCEP et al., 2015). The simulation data and codes needed to replicate all the figures and results in this paper are available at Dryad: <https://doi.org/10.5061/dryad.t1g1jwbtg>.

References

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023a). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970), 533–538. <https://doi.org/10.1038/s41586-023-06185-3>
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023b). Pangu-weather: Official repository [Software]. *Zenodo*. <https://doi.org/10.5281/zenodo.7678849>
- Bian, G.-F., Nie, G.-Z., & Qiu, X. (2021). How well is outer tropical cyclone size represented in the ERA5 reanalysis dataset? *Atmospheric Research*, 249, 105339. <https://doi.org/10.1016/j.atmosres.2020.105339>
- Bloemendaal, N., Haigh, I. D., de Moel, H., Muis, S., Haarsma, R. J., & Aerts, J. C. (2020). Generation of a global synthetic tropical cyclone hazard dataset using storm. *Scientific Data*, 7(1), 40. <https://doi.org/10.1038/s41597-020-0381-2>
- Bonavita, M. (2023). On some limitations of data-driven weather forecasting models. arXiv e-prints. arXiv:2309.
- Bretherton, C. S., Smith, C., & Wallace, J. M. (1992). An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, 5(6), 541–560. [https://doi.org/10.1175/1520-0442\(1992\)005<0541:aiomff>2.0.co;2](https://doi.org/10.1175/1520-0442(1992)005<0541:aiomff>2.0.co;2)
- Chavas, D. R., Lin, N., & Emanuel, K. (2015). A model for the complete radial structure of the tropical cyclone wind field. Part I: Comparison with observed structure. *Journal of the Atmospheric Sciences*, 72(9), 3647–3662. <https://doi.org/10.1175/jas-d-15-0014.1>
- Chavas, D. R., & Vigh, J. (2014). QSCAT-R: The QuikSCAT tropical cyclone radial structure dataset. *NCAR Tech. Note TN-5131STR*.

Acknowledgments

RJ acknowledge the generous support from the Katharine McCormick Fellowship. RJ and EB acknowledge support for the research from the NIH Grant R01HD104835. KE's research is part of the MIT Climate Grand Challenge on Weather and Climate Extremes, supported by the generosity of Eric and Wendy Schmidt through the recommendation of Schmidt Futures as part of its Virtual Earth System Research Institute (VESRI).

- Chowdhury, R. R., Kumar, S. P., & Chakraborty, A. (2022). Simultaneous occurrence of tropical cyclones in the northern Indian ocean: Differential response and triggering mechanisms. *Biogeochemical and Ecological Responses to Wind-or Tide-Induced Disturbances over Marginal Seas*.
- Emanuel, K. (1988). The maximum intensity of hurricanes. *Journal of the Atmospheric Sciences*, 45(7), 1143–1155. [https://doi.org/10.1175/1520-0469\(1988\)045<1143:tmioh>2.0.co;2](https://doi.org/10.1175/1520-0469(1988)045<1143:tmioh>2.0.co;2)
- Emanuel, K. (2017). Assessing the present and future probability of hurricane harvey's rainfall. *Proceedings of the National Academy of Sciences*, 114(48), 12681–12684. <https://doi.org/10.1073/pnas.1716222114>
- Emanuel, K., DesAutels, C., Holloway, C., & Korty, R. (2004). Environmental control of tropical cyclone intensity. *Journal of the Atmospheric Sciences*, 61(7), 843–858. [https://doi.org/10.1175/1520-0469\(2004\)061<0843:ecotci>2.0.co;2](https://doi.org/10.1175/1520-0469(2004)061<0843:ecotci>2.0.co;2)
- Emanuel, K., Ravela, S., Vivant, E., & Risi, C. (2006). A statistical deterministic approach to hurricane risk assessment. *Bulletin of the American Meteorological Society*, 87(3), 299–314. <https://doi.org/10.1175/bams-87-3-299>
- Emanuel, K., Sundararajan, R., & Williams, J. (2008). Hurricanes and global warming: Results from downscaling IPCC AR4 simulations. *Bulletin of the American Meteorological Society*, 89(3), 347–368. <https://doi.org/10.1175/bams-89-3-347>
- Gori, A., Lin, N., Schenkel, B., & Chavas, D. (2023). North Atlantic tropical cyclone size and storm surge reconstructions from 1950-present. *Journal of Geophysical Research: Atmospheres*, 128(5), e2022JD037312. <https://doi.org/10.1029/2022jd037312>
- Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., et al. (2016). High resolution model intercomparison project (HighResMIP v1. 0) for CMIP6. *Geoscientific Model Development*, 9(11), 4185–4208. <https://doi.org/10.5194/gmd-9-4185-2016>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Huang, M., Wang, Q., Jing, R., Lou, W., Hong, Y., & Wang, L. (2022a). Tropical cyclone full track simulation in the western north pacific based on random forests. *Journal of Wind Engineering and Industrial Aerodynamics*, 228, 105119. <https://doi.org/10.1016/j.jweia.2022.105119>
- Huang, M., Wang, Q., Li, Q., Jing, R., Lin, N., & Wang, L. (2021). Typhoon wind hazard estimation by full-track simulation with various wind intensity models. *Journal of Wind Engineering and Industrial Aerodynamics*, 218, 104792. <https://doi.org/10.1016/j.jweia.2021.104792>
- Huang, M., Wang, Q., Liu, M., Lin, N., Wang, Y., Jing, R., et al. (2022b). Increasing typhoon impact and economic losses due to anthropogenic warming in southeast China. *Scientific Reports*, 12(1), 14048. <https://doi.org/10.1038/s41598-022-17323-8>
- Irish, J. L., Resio, D. T., & Rattcliff, J. J. (2008). The influence of storm size on hurricane surge. *Journal of Physical Oceanography*, 38(9), 2003–2013. <https://doi.org/10.1175/2008jpo3727.1>
- James, M., & Mason, L. (2005). Synthetic tropical cyclone database. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 131(4), 181–192. [https://doi.org/10.1061/\(asce\)0733-950x\(2005\)131:4\(181\)](https://doi.org/10.1061/(asce)0733-950x(2005)131:4(181))
- Jing, R., Heft-Neal, S., Chavas, D. R., Griswold, M., Wang, Z., Clark-Ginsberg, A., et al. (2024). Global population profile of tropical cyclone exposure from 2002 to 2019. *Nature*, 626(7999), 549–554. <https://doi.org/10.1038/s41586-023-06963-z>
- Jing, R., & Lin, N. (2019). Tropical cyclone intensity evolution modeled as a dependent hidden Markov process. *Journal of Climate*, 32(22), 7837–7855. <https://doi.org/10.1175/jcli-d-19-0027.1>
- Jing, R., & Lin, N. (2020). An environment-dependent probabilistic tropical cyclone model. *Journal of Advances in Modeling Earth Systems*, 12(3), e2019MS001975. <https://doi.org/10.1029/2019ms001975>
- Jing, R., Lin, N., Emanuel, K., Vecchi, G., & Knutson, T. R. (2021). A comparison of tropical cyclone projections in a high-resolution global climate model and from downscaling by statistical and statistical-deterministic methods. *Journal of Climate*, 34(23), 9349–9364.
- Kareem, A., Hu, L., Guo, Y., & Kwon, D.-K. (2019). Generalized wind loading chain: Time-frequency modeling framework for nonstationary wind effects on structures. *Journal of Structural Engineering*, 145(10), 04019092. [https://doi.org/10.1061/\(asce\)st.1943-541x.0002376](https://doi.org/10.1061/(asce)st.1943-541x.0002376)
- Klotzbach, P. J., Chavas, D. R., Bell, M. M., Bowen, S. G., Gibney, E. J., & Schreck, C. J., III. (2022). Characterizing continental us hurricane risk: Which intensity metric is best? *Journal of Geophysical Research: Atmospheres*, 127(18), e2022JD037030. <https://doi.org/10.1029/2022jd037030>
- Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., & Schreck, C. J. (2018). International best track archive for climate stewardship (IBTrACS) project, version 4.v04r00. <https://doi.org/10.25921/82ty-9e16>
- Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data [Dataset]. *Bulletin of the American Meteorological Society*, 91(3), 363–376. <https://doi.org/10.1175/2009bams2755.1>
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., et al. (2022). Graphcast: Learning skillful medium-range global weather forecasting. arXiv preprint arXiv:2212.12794.
- Lee, C.-Y., Camargo, S. J., Sobel, A. H., & Tippett, M. K. (2020). Statistical–dynamical downscaling projections of tropical cyclone activity in a warming climate: Two diverging genesis scenarios. *Journal of Climate*, 33(11), 4815–4834. <https://doi.org/10.1175/jcli-d-19-0452.1>
- Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2016). Rapid intensification and the bimodal distribution of tropical cyclone intensity. *Nature Communications*, 7(1), 10625. <https://doi.org/10.1038/ncomms10625>
- Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2018). An environmentally forced tropical cyclone hazard model. *Journal of Advances in Modeling Earth Systems*, 10(1), 223–241. <https://doi.org/10.1002/2017ms001186>
- Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., et al. (2023). Climate change 2023: Synthesis report. Contribution of working groups I, II and III to the sixth assessment report of the intergovernmental panel on climate change.
- Lin, N., & Chavas, D. (2012). On hurricane parametric wind and applications in storm surge modeling. *Journal of Geophysical Research*, 117(D9), D09120. <https://doi.org/10.1029/2011jd017126>
- Lin, N., Emanuel, K., Oppenheimer, M., & Vanmarcke, E. (2012). Physically based assessment of hurricane surge threat under climate change. *Nature Climate Change*, 2(6), 462–467. <https://doi.org/10.1038/nclimate1389>
- Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.
- Marsooli, R., Lin, N., Emanuel, K., & Feng, K. (2019). Climate change exacerbates hurricane flood hazards along us Atlantic and gulf coasts in spatially varying patterns. *Nature Communications*, 10(1), 3785. <https://doi.org/10.1038/s41467-019-11755-z>
- Meiler, S., Vogt, T., Bloemendaal, N., Ciullo, A., Lee, C.-Y., Camargo, S. J., et al. (2022). Intercomparison of regional loss estimates from global synthetic tropical cyclone models. *Nature Communications*, 13(1), 6156. <https://doi.org/10.1038/s41467-022-33918-1>
- Mendelsohn, R., Emanuel, K., Chonabayashi, S., & Bakkensen, L. (2012). The impact of climate change on global tropical cyclone damage. *Nature Climate Change*, 2(3), 205–209. <https://doi.org/10.1038/nclimate1357>
- Nam, H., Lee, H., Park, J., Yoon, W., & Yoo, D. (2021). Reducing domain gap by reducing style bias. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8690–8699).
- Nandi, A., Mandal, A., Wilson, M., & Smith, D. (2016). Flood hazard mapping in Jamaica using principal component analysis and logistic regression. *Environmental Earth Sciences*, 75(6), 1–16. <https://doi.org/10.1007/s12665-016-5323-0>

- NCEP, Service, N. W., NOAA, & of Commerce, U. D. (2015). NCEP GFS 0.25 degree global forecast grids historical archive [Dataset]. *Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory*. <https://rda.ucar.edu/datasets/ds084.1/>
- Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., et al. (2022). Fourcastnet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv preprint arXiv:2202.11214.
- Pérez, P., Gangnet, M., & Blake, A. (2003). *Poisson image editing*. ACM SIGGRAPH 2003 Papers.
- Pörtner, H.-O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., et al. (2022). *Climate change 2022: Impacts, adaptation and vulnerability*. IPCC Sixth Assessment Report.
- Powell, M. D., & Reinhold, T. A. (2007). Tropical cyclone destructive potential by integrated kinetic energy. *Bulletin of the American Meteorological Society*, 88(4), 513–526. <https://doi.org/10.1175/bams-88-4-513>
- Ruan, Z., & Wu, Q. (2022). Relationship between size and intensity in north Atlantic tropical cyclones with steady radii of maximum wind. *Geophysical Research Letters*, 49(3), e2021GL095632. <https://doi.org/10.1029/2021gl095632>
- Schenkkel, B. A., Lin, N., Chavas, D., Oppenheimer, M., & Brammer, A. (2017). Evaluating outer tropical cyclone size in reanalysis datasets using QuikSCAT data. *Journal of Climate*, 30(21), 8745–8762. <https://doi.org/10.1175/jcli-d-17-0122.1>
- Tippett, M. K., Camargo, S. J., & Sobel, A. H. (2011). A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *Journal of Climate*, 24(9), 2335–2357. <https://doi.org/10.1175/2010jcli3811.1>
- Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., et al. (2018). Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 969–977).
- Ullrich, P. A., & Zarzycki, C. M. (2017). Tempestextremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, 10(3), 1069–1090. <https://doi.org/10.5194/gmd-10-1069-2017>
- Velden, C., Harper, B., Wells, F., Beven, J. L., Zehr, R., Olander, T., et al. (2006). The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bulletin of the American Meteorological Society*, 87(9), 1195–1210. <https://doi.org/10.1175/bams-87-9-1195>
- Vickery, P., Skerlj, P., & Twisdale, L. (2000). Simulation of hurricane risk in the us using empirical track model. *Journal of structural engineering*, 126(10), 1222–1237. [https://doi.org/10.1061/\(asce\)0733-9445\(2000\)126:10\(1222\)](https://doi.org/10.1061/(asce)0733-9445(2000)126:10(1222))
- Wang, S., Lin, N., & Gori, A. (2022). Investigation of tropical cyclone wind models with application to storm tide simulations. *Journal of Geophysical Research: Atmospheres*, 127(17), e2021JD036359. <https://doi.org/10.1029/2021jd036359>
- Wei, L., Zhang, S., Gao, W., & Tian, Q. (2018). Person transfer GAN to bridge domain gap for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 79–88).
- Zarzycki, C. M., Ullrich, P. A., & Reed, K. A. (2021). Metrics for evaluating tropical cyclones in climate data. *Journal of Applied Meteorology and Climatology*, 60(5), 643–660. <https://doi.org/10.1175/jamc-d-20-0149.1>
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer vision—ECCV 2014: 13th European conference, Zurich, Switzerland, September 6–12, 2014, proceedings, part I 13* (pp. 818–833).