

# TC-GEN: Data-driven Tropical Cyclone Downscaling using Machine Learning-Based High-resolution Weather Model

Renzhi Jing<sup>1</sup>, Jianxiong Gao<sup>2</sup>, Yunuo Cai<sup>2</sup>, Dazhi Xi<sup>3</sup>, Yinda Zhang<sup>4</sup>, Yanwei Fu<sup>2</sup>, Kerry A. Emanuel<sup>5</sup>, Noah S. Diffenbaugh<sup>1</sup>, and Eran Bendavid<sup>1</sup>

<sup>1</sup>Stanford University

<sup>2</sup>Fudan University

<sup>3</sup>Princeton University

<sup>4</sup>Google LLC

<sup>5</sup>Massachusetts Institute of Technology

January 16, 2024

## 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 introduce 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 is able to 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 integrate storm seeds into the surrounding while maintain the seed structure, and a recurrent neural network-based model for correcting the biases in maximum wind speed. 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. Taking advantage of the computational efficiency of ML-GWM, our approach shows substantial potential for TC regional hazard and risk assessment.

# TC-GEN: Data-driven Tropical Cyclone Downscaling using Machine Learning-Based High-resolution Weather Model

Renzhi Jing<sup>1,2</sup>, Jianxiong Gao<sup>3</sup>, Yunuo Cai<sup>4</sup>, Dazhi Xi<sup>5</sup>, Yinda Zhang<sup>6</sup>,  
Yanwei Fu<sup>4</sup>, Kerry Emanuel<sup>7</sup>, Noah S. Diffenbaugh<sup>1,8,9</sup>, Eran Bendavid<sup>2,10</sup>

<sup>1</sup>Woods Institute for the Environment, Stanford University, Stanford, CA, USA

<sup>2</sup>School of Medicine, Stanford University, Stanford, CA, USA

<sup>3</sup>Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai, China

<sup>4</sup>School of Data Science and MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China

<sup>5</sup>Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA

<sup>6</sup>Google LLC, 1600 Amphitheatre Pkwy, Mountain View, CA, USA

<sup>7</sup>Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology,

Cambridge, MA, USA

<sup>8</sup>Department of Earth System Science, Stanford University, Stanford, CA, USA

<sup>9</sup>Doerr School of Sustainability, Stanford University, Stanford, CA, USA

<sup>10</sup>Department of Health Policy, Stanford University, Stanford, CA, USA

## Key Points:

- We develop a novel approach for tropical cyclone downscaling using a machine learning-based high-resolution global weather model.
- We generate and integrate synthetic tropical cyclone seeds into the surrounding environment using a data-driven approach.
- The new tropical cyclone downscaling approach is capable of simulating the two-way interactions between storms and the environment as a unified system.

---

Corresponding author: Renzhi Jing, [rjing@stanford.edu](mailto:rjing@stanford.edu)

Corresponding author: Yanwei Fu, [yanweifu@fudan.edu.cn](mailto:yanweifu@fudan.edu.cn)

## 24 Abstract

25 Synthetic downscaling of tropical cyclones (TCs) is critically important to estimate the  
 26 long-term hazard of rare high-impact storm events. Existing downscaling approaches rely  
 27 on statistical or statistical-deterministic models that are capable of generating large samples  
 28 of synthetic storms with characteristics similar to observed storms. However, these models  
 29 do not capture the complex two-way interactions between a storm and its environment. In  
 30 addition, these approaches either necessitate a separate TC size model to simulate storm  
 31 size or involve post-processing to introduce asymmetries in the simulated surface wind. In  
 32 this study, we present an innovative data-driven approach for TC synthetic downscaling.  
 33 Using a machine learning-based high-resolution global weather model (ML-GWM), our ap-  
 34 proach is able to simulate the full life cycle of a storm with asymmetric surface wind that  
 35 accounts for the two-way interactions between the storm and its environment. This ap-  
 36 proach consists of multiple components: a data-driven model for generating synthetic TC  
 37 seeds, a blending method that seamlessly integrate storm seeds into the surrounding while  
 38 maintain the seed structure, and a recurrent neural network-based model for correcting the  
 39 biases in maximum wind speed. Compared to observations and synthetic storms simulated  
 40 using existing statistical-deterministic and statistical downscaling approaches, our method  
 41 shows the ability to effectively capture many aspects of TC statistics, including track den-  
 42 sity, landfall frequency, landfall intensity, and outermost wind extent. Taking advantage of  
 43 the computational efficiency of ML-GWM, our approach shows substantial potential for TC  
 44 regional hazard and risk assessment.

## 45 Plain Language Summary

46 Tropical cyclones (TCs) cause significant destruction each year. It is crucial to accu-  
 47 rately assess the risks they present, but this is challenging due to a scarcity of historical  
 48 data. A commonly used approach involves creating a large number of synthetic TCs that  
 49 share key characteristics with real storms, enabling an effective regional risk assessment.  
 50 However, traditional synthetic TC generation approaches do not capture the complex inter-  
 51 actions between storms and their larger-scale environment. Furthermore, these approaches  
 52 do not adequately represent the asymmetric structure of TCs, despite the crucial role that  
 53 they play in modeling storm-related hazards such as rainfall and surges. Recently, advances  
 54 in machine learning-based global weather forecasting (ML-GWM) have provided highly ac-  
 55 curate and efficient high-resolution global weather forecasts that surpass the capabilities of  
 56 conventional numerical weather forecasting. In this study, we introduce a novel synthetic TC  
 57 generation approach, which we call the synthetic TC-GENerative Model (or "TC-GEN"),  
 58 leveraging the state-of-the-art ML-GWM. We show that TC-GEN can generate a large num-  
 59 ber of synthetic storms that allow the interaction between the storm and its environment.  
 60 We evaluate the performance of TC-GEN in various aspects, including several landfall char-  
 61 acteristics, which are of the most importance for local TC risk analysis. Our study also  
 62 serves as a compelling example of the transformative impact of machine learning and data  
 63 science in revolutionizing climate studies during the era of artificial intelligence.

## 64 1 Introduction

65 Tropical cyclones (TCs) are among the most destructive natural disasters, causing  
 66 substantial damage and losses in multiple ocean basins annually. In a warming climate, it is  
 67 projected that TCs are likely to become more intense, with an expected increase in both the  
 68 peak maximum wind speed and the proportion of strong TCs in the future (Pörtner et al.,  
 69 2022; H. Lee et al., 2023). Accurate assessment of TC tracks and intensities is fundamental to  
 70 reducing the impacts of landfalling storms. However, with around 90 storms occurring every  
 71 year and an average of only 20 making landfall, this task is challenging due to the shortage of  
 72 historical data required for regional risk assessment. To overcome data deficiency, a widely  
 73 used approach is to generate synthetic TCs that are capable of responding to various climate

74 conditions (K. Emanuel et al., 2008; C.-Y. Lee et al., 2018; Jing & Lin, 2020). Using large  
75 samples of synthetic storms, including extreme events with extended return periods, enables  
76 a comprehensive risk assessment for specific regions in both current and future climates.

77 Previous studies on synthetic TC generation have primarily employed two main ap-  
78 proaches: (1) statistical re-sampling, and (2) physical-based downscaling methods. Statisti-  
79 cal re-sampling models simulate TC genesis, tracks, and intensities (maximum wind speed  
80 or minimum central pressure) based purely on historical observational datasets, without  
81 considering environmental conditions. Examples of such models include those developed by  
82 (Vickery et al., 2000; James & Mason, 2005; Bloemendaal et al., 2020). These models typ-  
83 ically require a limited number of input variables, have low computational costs, and thus  
84 are easily applicable on a global scale. However, these models are not based on physical  
85 principles and cannot be accurately applied to a non-stationary climate due to changes in  
86 the background environment. On the other hand, physically-based downscaling methods  
87 relate TC characteristics to the large-scale background environmental conditions, making  
88 them environment-dependent. These methods can be statistical-deterministic (e.g., models  
89 developed by (K. Emanuel et al., 2006)) or purely statistical (e.g., models developed by  
90 (C.-Y. Lee et al., 2018; Jing & Lin, 2020)). Such environment-dependent approaches are  
91 capable of simulating the TC climatology in future climate scenarios and, therefore, are suit-  
92 able for climate change studies (K. Emanuel et al., 2008; C.-Y. Lee et al., 2020; Jing et al.,  
93 2021). Since the first synthetic TC downscaling approach in this family of models appeared  
94 in 2006, significant advances have been made in each of the three components (Huang et  
95 al., 2021; Huang, Wang, Jing, et al., 2022), and these approaches have been widely used  
96 for applications such as TC-induced surge risk assessment (Lin et al., 2012), regional loss  
97 assessment (Meiler et al., 2022; Huang, Wang, Liu, et al., 2022), and TC-induced wind load  
98 analysis (Kareem et al., 2019).

99 Both statistical and statistical-deterministic synthetic downscaling methods simulate  
100 the complete lifecycle of TCs using environmental parameters derived from the background  
101 environment. However, they cannot simultaneously simulate the two-way interactions be-  
102 tween the storm and its surrounding environment; therefore, the environment cannot re-  
103 spond correspondingly to the development of the storm. Furthermore, traditional ap-  
104 proaches do not comprehensively simulate the characteristics of TCs (i.e. genesis, track,  
105 intensity, and size) as a cohesive system. For example, although the TC intensity is deter-  
106 mined based on environmental predictors along the track, the storm track is predominantly  
107 driven by background winds, which is independent of the intensity component. Given the  
108 clear and strong correlation between these components, it is prudent to consider the poten-  
109 tial correlations between these components (Ruan & Wu, 2022). Moreover, the asymmetries  
110 in the TC surface wind field are not directly captured. Some approaches do not output the  
111 TC size and require a separate size component to determine the outer radius of the storm  
112 by random sampling from historical data (Jing & Lin, 2020). Other methods provide the  
113 radius of maximum wind (K. Emanuel et al., 2008); however, they require an additional  
114 parametric wind model to generate the full surface wind field, followed by post-processing  
115 to incorporate asymmetries related to storm translation speed and wind shear (Lin et al.,  
116 2012; Lin & Chavas, 2012).

117 The ideal synthetic downscaling method would simultaneously simulate all character-  
118 istics of the TC as an integrated system, including the interactions between storms and the  
119 surrounding environment, to generate detailed wind fields with greater accuracy but similar  
120 computational efficiency to that of traditional statistical and statistical-deterministic down-  
121 scaling methods. Recent progress in machine learning-based high-resolution global weather  
122 modeling (ML-GWM) (Pathak et al., 2022; Bi et al., 2022; Lam et al., 2022) has made  
123 this possible. ML-GWM systems are based on three-dimensional neural networks that are  
124 trained on high-quality reanalysis datasets, such as the ERA5 reanalysis (Hersbach et al.,  
125 2020), to predict weather around the globe. A significant advantage of ML-GWMs is their  
126 substantially lower computational costs compared to traditional numerical weather fore-

127 casting, while still operating at high spatial resolutions. Representing the cutting edge of  
128 ML-GWM, Pangu-Weather (Bi et al., 2022) has outperformed the operational Integrated  
129 Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts  
130 in medium-range forecasting, with speeds more than 10,000 times faster. The high spa-  
131 tial resolution of 0.25 degrees also enables Pangu-Weather to precisely track TCs based on  
132 simulation results.

133 In this study, we leverage ML-GWM to create a novel ML-based approach for synthetic  
134 TCs downscaling, which we call the synthetic TC-GENerative Model (or "TC-GEN"). This  
135 involves generating a synthetic TC seed for each storm through a data-driven process, merg-  
136 ing it with the background environment, and simulating both the storm and its surround-  
137 ings simultaneously with Pangu-Weather. To achieve this, we first determine the annual  
138 frequency, date, and location of synthetic TCs using an existing environment-based TC  
139 genesis model. Next, we perform a Principal Component Analysis on all historical TCs at  
140 genesis, identifying the principal components that effectively capture most of the variances  
141 in TC genesis. Using these principal components, we generate synthetic TC seeds with  
142 weights derived from historical data. We then integrate these TC seeds into the surround-  
143 ing environment using Poisson blending, a technique widely used in image processing to  
144 seamlessly merge two images, ensuring that the TC seeds are naturally embedded within  
145 the larger environmental context while still maintaining important wind structures. Finally,  
146 we run Pangu-Weather using their pre-trained model, which enables the joint simulation  
147 of both the storm and its surrounding environment, and bias-correct simulated intensity to  
148 real intensity. This integrated approach allows us to gather key characteristics of the TC,  
149 such as the track and the maximum wind speed. It also provides spatial details such as  
150 the full wind field, allowing for a direct derivation of the outermost extent of the storm.  
151 It is worth noting that several key steps of this ML-based method are data driven, relying  
152 heavily on historical data that lack substantial input from physics. Furthermore, while the  
153 spatial resolution of 0.25 ° is a high resolution for global climate models, it is still too coarse  
154 to accurately resolve the inner core and the structure of a storm. Therefore, we should use  
155 the simulated three-dimensional structure including the horizontal surface wind field with  
156 care, as it is likely to be unrealistic.

157 Despite these limitations, our ML-based method offers a unique set of collective ad-  
158 vantages compared to previous TC downscaling approaches: 1) Holistic simulation: unlike  
159 previous studies where the genesis, track, and intensity of the storm are simulated sep-  
160 arately, this approach can simulate these three storm components together as a cohesive  
161 system; 2) Integrated simulation: similar to numerical modeling, this approach can sim-  
162 ulate both the storm and its environment simultaneously, thus allowing for the two-way  
163 interactions between the storm and the environment; 3) Intrinsic asymmetry: while the TC  
164 core and intensity may not be fully resolved, this approach has the capability to provide  
165 crucial asymmetric characteristics in the TC surface wind field. This, in turn, allows for the  
166 inference of the asymmetric outermost wind extent of a storm, which is essential for studying  
167 tropical cyclone-induced hazards such as surges and heavy rainfall; 4) Efficiency: utilizing  
168 pre-trained Pangu-Weather, this approach inherits the efficiency of statistical downscaling  
169 methods that require little computational resources, enabling large samples of synthetic TCs  
170 to be generated in days, a time frame comparable to the work of (Bloemendaal et al., 2020);  
171 5) Scalability: this approach can be easily applied to other ocean basins and other high-  
172 resolution climate datasets, including future climate projections such as those in CMIP6. A  
173 successful extension of this approach is achieved when the climate dataset used for training  
174 processes high resolution for detecting storm eyes, provides a reasonable representation of  
175 the storm's outer size, and involves the development of a corresponding pre-trained machine  
176 learning model as the core simulator.

177 To evaluate TC-GEN, we generate a large sample of synthetic TCs and compare those  
178 simulated storms with historical observations. We also place TC-GEN in context with  
179 an existing statistical-deterministic downscaling approach, represented by Emaneul et al.

180 (2008) (KE08, (K. Emanuel et al., 2008)) and a statistical downscaling method, represented  
 181 by the Princeton environment-dependent probabilistic model (PepC, (Jing & Lin, 2020)).  
 182 We choose these two existing approaches as they are both environment dependent and  
 183 have distinct model components. The metrics we use for comparison include the density of  
 184 the tracks over the ocean, the maximum lifetime intensity, the landfall frequency, and the  
 185 landfall intensity distributions. Given that KE08 and PepC lack the ability to simulate the  
 186 outer size of TC, we only compare the distribution of TC outer size simulated by TC-GEN  
 187 with that of the historical TCs identified using reanalysis data. In all of these metrics,  
 188 we demonstrate a strong alignment between simulated storms and observational data. We  
 189 further assess the adaptability of TC-GEN to different reanalysis datasets through two  
 190 analyses: one with the ERA5 reanalysis, on which Pangu-Weather is trained, but using  
 191 different temporal resolutions for model initialization, and the other using an alternative  
 192 reanalysis dataset from the National Centers for Environmental Prediction (NCEP). We  
 193 show that the effectiveness of TC-GEN depends on the consistency between the training  
 194 reanalysis dataset used to train the ML-GMW and the data used for model initialization.  
 195 Based on these analyses, we summarize the strength and limitations of TC-GEN and propose  
 196 potential improvements for future work.

## 197 **2 Pre-trained Models and Data**

### 198 **2.1 Neural network-based global weather model**

199 In this study, we use Pangu-Weather as the core ML-GWM to generate synthetic TCs.  
 200 Pangu-weather is an artificial intelligence-based model for medium-range global weather  
 201 forecasting, which has been shown to outperform the operational integrated forecasting  
 202 system of the European Center for Medium-Range Weather Forecasts (ECMWF) (Bi et al.,  
 203 2022).

204 Pangu-Weather is trained on atmospheric reanalysis data from the ERA5 reanalysis  
 205 data (Hersbach et al., 2020), using 69 atmospheric and surface variables as input and fore-  
 206 casting these variables for the subsequent time step. The 69 variables include 65 upper-air  
 207 variables (geopotential height, specific humidity, temperature, u and v component of wind at  
 208 13 pressure levels) plus four surface weather variables (2m temperature, u- and v- component  
 209 of 10m wind speed, and mean sea level pressure). Pangu-Weather offers four pre-trained  
 210 models that are capable of forecasting global weather data 1, 3, 6, and 24 hours ahead. We  
 211 use the model with a lead time of 6 hours since it is consistent with the temporal resolution  
 212 of TC IBTrACS data (see Section 2.2). Pangu-Weather has the ability to produce accurate  
 213 simulations that span multiple consecutive days, a time frame that is adequate to simulate  
 214 the entire life span of the most TCs. Pangu-Weather is open-source, and the pre-trained  
 215 models can be downloaded at <https://github.com/198808xc/Pangu-Weather>.

216 Pangu-Weather has several distinct advantages, making it highly suitable for this study.  
 217 First, Pangu-Weather is among the state-of-the-art ML-GWM systems with the highest  
 218 performance in weather forecasting. Second, Pangu-Weather has a high horizontal resolution  
 219 of  $0.25^\circ \times 0.25^\circ$ . Although the thermodynamic processes of the TCs cannot be fully resolved  
 220 at this spatial resolution, it is sufficient to detect the low pressure center, which enables  
 221 effective storm eye tracking. Third, Pangu-Weather is computationally efficient (more than  
 222 10,000 times faster than operational dynamical models), making it capable of simulating  
 223 large numbers of synthetic storms. In our experiments, we run Pangu-Weather with 8 RTX  
 224 A6000 GPUs, which adequately support all our computational tasks.

### 225 **2.2 Tropical Cyclone Data**

226 TC observations are extracted from IBTrACS data (version v04r00) (Knapp et al.,  
 227 2018, 2010), which includes the location (latitude and longitude) of each storm every 6  
 228 hours, along with its maximum sustained wind speeds measured at a height of 10 meters

229 above the sea surface. We use historical TC data in the North Atlantic Basin between 1979  
 230 and 2022 to generate TC seeds, correct biases in intensities simulated from Pangu-Weather,  
 231 and evaluate TC-GEN performance. We use extended TC data dating back to 1900 to  
 232 analyze landfall frequency, which provides a more precise assessment of sampling errors  
 233 with a larger number of TC tracks.

### 234 **2.3 Atmospheric Reanalysis Data**

235 Pangu-Weather requires global atmospheric and surface variables at the current time  
 236 step to forecast the next step. To be consistent with the input setup for the Pangu-Weather  
 237 model, we use ERA5 reanalysis with a resolution of  $0.25^\circ \times 0.25^\circ$ , the highest available  
 238 spatial resolution, and a temporal resolution of both 6-hour and monthly reanalysis data  
 239 to initiate Pangu-Weather. Using 6-hour reanalysis data for Pangu-Weather initialization  
 240 is intuitive, as it ensures alignment with the model setup and prevents domain gaps. How-  
 241 ever, generating synthetic TCs with 6-hourly reanalysis data presents two challenges. First,  
 242 introducing a synthetic TC to the 6-hourly reanalysis data may result in multiple storms  
 243 developing on the same map. This does not reflect reality, as the simultaneous occurrence  
 244 of multiple storms is relatively rare (Chowdhury et al., 2022). Second, generating synthetic  
 245 TCs by initiating Pangu-Weather with 6-hourly data requires downloading extensive his-  
 246 torical ERA5 reanalysis data, resulting in substantial data storage requirements. On the  
 247 contrary, starting Pangu-Weather with monthly data significantly reduces the data storage  
 248 burden and mitigates the problem of simultaneous occurrence. In this study, we use both  
 249 6-hourly and monthly reanalysis data for Pangu-Weather initialization, and under both sce-  
 250 narios we use the pre-trained model with a lead time of 6 hours to simulate storms. We  
 251 denote these two scenarios as 'TC-GEN hourly' and 'TC-GEN monthly' in the following  
 252 text and results.

253 In addition to ERA5 reanalysis, we test TC-GEN performance when initiating Pangu-  
 254 Weather with the NCEP GFS from Global Forecast Grids Historical Archive at  $0.25^\circ \times 0.25^\circ$   
 255 resolution (for Environmental Prediction/National Weather Service/NOAA/US Depart-  
 256 ment of Commerce, 2015). We opt for this reanalysis dataset because it provides all 69  
 257 variables required by Pangu-Weather, and it is available at a resolution of  $0.25^\circ \times 0.25^\circ$  with  
 258 global coverage.

### 259 **2.4 Existing downscaling approaches**

260 We contextualize TC-GEN by comparing it with two existing synthetic downscaling  
 261 approaches. The first is the statistical-deterministic model developed by Emanuel et al.  
 262 (2008) (K. Emanuel et al., 2008), which applies a random seeding method to initiate the  
 263 storm, a beta and advection model based on local winds to propagate the storm, and a  
 264 deterministic Coupled Hurricane Intensity Prediction System (CHIPS; (K. Emanuel et al.,  
 265 2004)) model to estimate the intensity of the storm based on the local thermodynamic state  
 266 of the atmosphere and ocean. The model has been extensively applied to assess TC hazards  
 267 (K. Emanuel, 2017; Marsooli et al., 2019), economic losses (Mendelsohn et al., 2012; Meiler  
 268 et al., 2022), and changes in TC climatology under future projected climate conditions  
 269 (K. Emanuel et al., 2008; Jing et al., 2021). Here, we compare our ML-based TC-GEN  
 270 results with a total of 4,100 synthetic TCs from 1980-2020 that intensify and reach TC  
 271 strength (lifetime maximum intensity exceeds 34 kt).

272 We also compare our TC-GEN results with the "PepC", a statistical synthetic down-  
 273 scaling approach of Jing and Lin (2020) (Jing & Lin, 2019). PepC has three components, a  
 274 genesis model, a track model, and an intensity model. The genesis component determines  
 275 annual frequency, as well as the time and location of weak vortices. The genesis model is de-  
 276 veloped using Poisson regression based on four large-scale environmental variables: absolute  
 277 vorticity, wind shear, relative humidity, and maximum potential intensity (K. A. Emanuel,  
 278 1988), which are contained in a genesis index (Tippett et al., 2011). After initialization,

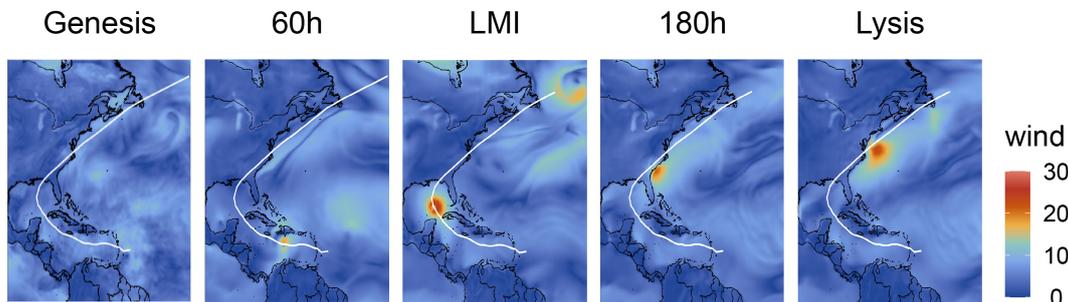
279 an analog-wind track model is used to propagate the storm based on analog factors (from  
 280 historical track patterns) and local in situ winds. The intensity of the storm is simulated as  
 281 a Markov process using an environment-dependent hurricane intensity model (Jing & Lin,  
 282 2019). In (Jing & Lin, 2020), the authors generated more than 55,000 TCs from 100 inde-  
 283 pendent realizations of the genesis of TCs during the period 1979-2014. To compare with  
 284 our ML-based TC-GEN results, we randomly select 10 independent 36-year realizations,  
 285 including a total of 3,690 TCs that intensify and reach TC strength.

### 286 **3 TC-GEN: data-driven synthetic TC downscaling approach**

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

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

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



**Figure 1. TC-GEN.** We propose an AI-empowered approach for synthetic TC downscaling where TC seed is created through a data-driven approach and then incorporated into a reanalysis environment map. Both the TC and its surrounding environment are then simulated in an interactive manner. The idea is illustrated in the figure, which presents five snapshots of total wind at different stages of the storm's life cycle.

#### 301 **3.1 Annual counts, date and locations of TC Seeds**

302 The first step of TC-GEN is to determine the annual frequency and the date and  
 303 location of the synthetic TCs. We use an environment-based hierarchical Poisson genesis  
 304 model to determine the number of synthetic storms that occur each year, as well as when  
 305 and where they originate over the ocean basin. This model is the genesis component of PepC  
 306 (Jing & Lin, 2020), which has been shown to generate TC formations that align closely with  
 307 observations, including interannual variations. This genesis model simulates the formation

of TCs for each month. When initializing Pangu-Weather with 6-hour reanalysis data, we randomly assign the precise date and time for each storm within the month of its genesis.

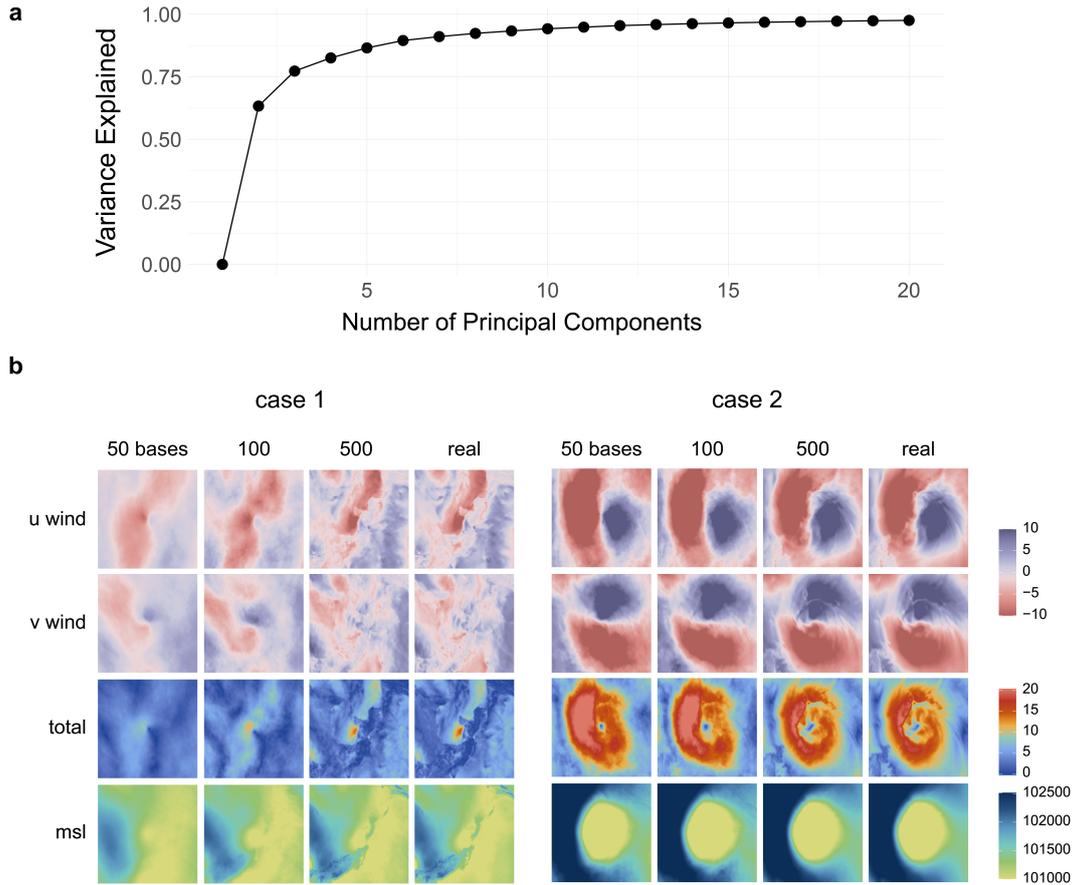
### 3.2 Structured TC seeds generation via data-driven approach

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 principal component analysis (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.

Principal Component Analysis (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 dataset. 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 (Nandi et al., 2016; Bretherton et al., 1992), and we refer the reader to the survey (Abdi & Williams, 2010) for more details.

We use 690 historical TCs in the period 1979 - 2022 to create 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 1600 kilometers), centered on the storm's eye. This circular region is large enough to encompass the outermost extent of most TCs at genesis. We collect TC seeds for each of the 69 atmospheric and surface variables from the ERA5 reanalysis for all  $M = 690$  storms, which form a matrix with dimensions of  $128 \times 128 \times 69 \times 690$ , representing historical TC seeds. Next, we perform PCA to reduce the dimensions of the data. The cumulative sum of the largest 20 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. To ensure quality, we use 500 principal components that represent the majority of the variations in actual environmental fields.

We create synthetic TC seeds by randomly sampling the low-dimensional space based on the top 500 principal components. 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. In Figure 3 we show three synthetic TCs at their genesis. We show four important variables (mean sea 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.

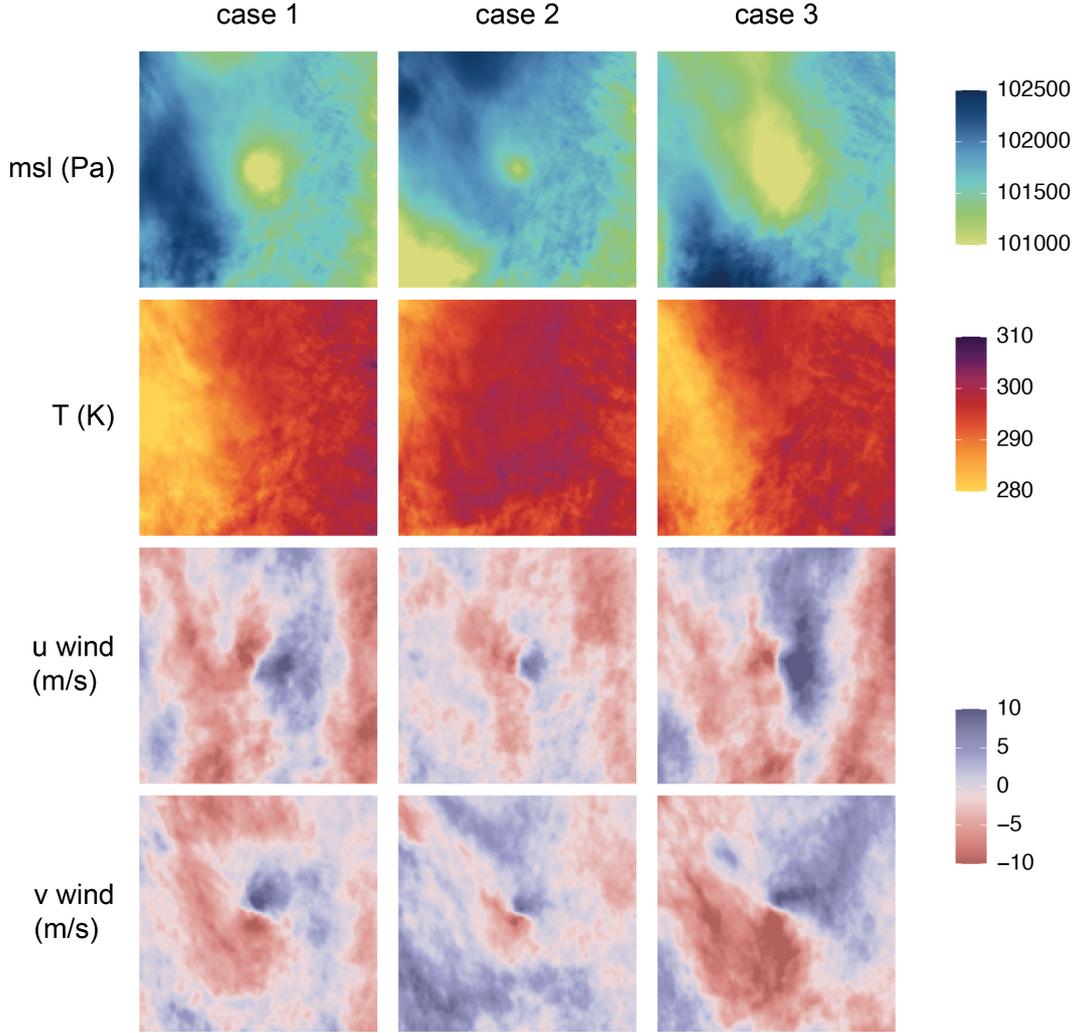


**Figure 2. Principal Component Analysis on TC genesis.** We run PCA on TC historical data to learn a low dimensional representation for TC seeds at genesis. Subplot (a) shows the cumulative sum of the eigenvalues sorted in descending order, where each point represents the proportion of total variance 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, using 50, 100, 500 principal components respectively. Both TC seeds are almost completely reconstructed using 500 principal components.

### 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 Sec 3.1). This step is critical for Pangu-Weather to simulate the environment along with the synthetic TC seeds. To achieve seamless integration of 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  $S \subset R^2$  (for example, color defined on pixel grids), and



**Figure 3. Synthetic TC Seeds.** This figure shows three synthetic TC seeds at genesis using the data-driven approach with 500 principal components. Each column shows one case, where mean sea level pressure, temperature, surface u- and v-winds are visualized.

$g$  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  $\Omega$ , such that

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

370 , where  $\nabla \cdot = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}]$ . This optimization aims to achieve a blended result where, within the  
 371 source region, it closely resembles the gradient of  $g$ , while at the boundary it should be  
 372 similar to  $f^*$ . For more details and numerical solutions, we refer the reader to (Pérez et al.,  
 373 2003).

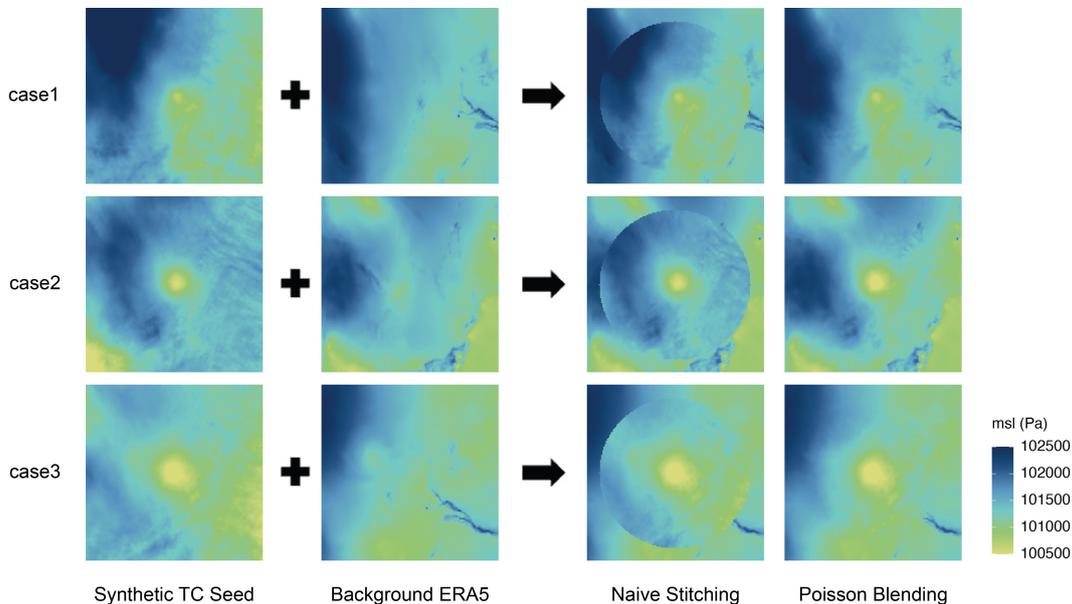
374 As an analogy to our problem,  $f^*$  and  $g$  represent the global environment map (that  
 375 is, target matrix) and synthetic TC seeds (that is, source matrix), respectively. In this  
 376 context, Poisson blending aims to combine these two matrices, ensuring that the gradients  
 377 of the blended results within the source region closely resemble the synthetic TC seeds, while  
 378 at the boundary, they should be similar to the background environment. In practice, we find  
 379 that naive Poisson blending still results in noticeable artifacts. Therefore, we adopt a linear

380 blending technique that incorporates both the source and target gradients, with a weight  
 381 determined by the distance to the boundary. The optimization is modified to accommodate  
 382 this approach:

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

383 where  $d$  is the distance of each pixel to the boundary and  $R$  represents the radius of the  
 384 blended region. This optimization guarantees a smooth integration of TC seeds into the  
 385 global environment map, with natural transitions at the boundaries and the structure of  
 386 TC well maintained.

387 We use this Poisson blending approach to seamlessly integrate the 69 atmospheric and  
 388 surface variables of synthetic TC seeds within a radius of 64 grid cells (approximately 1600  
 389 km) into the corresponding ERA5 reanalysis data. In Figure 4, we show the Poisson blend  
 390 process using mean sea level pressure as an example, which clearly reveals the eye of the  
 391 storm. Similarly, we show three cases as examples of weak, moderate, and strong storms,  
 392 illustrating the effectiveness of Poisson blending across a range of storm characteristics.



**Figure 4. 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 eye 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.

### 393 3.4 Simulation and tracking

394 After seamlessly integrating the generated TC seeds into the background environment  
 395 map, we proceed to run Pangu-Weather, simulating the entire life cycle of each storm over  
 396 a 15-day period, with 6-hour intervals between each step. To track the location of storm  
 397 center from Pangu-Weather outputs, we formulate a Gaussian kernel over mean sea level  
 398 pressure and vorticity at 10 meters, to identify the local maximum vorticity or minimum

399 pressure associated with the characteristic bell-shaped symmetric structure typical of a  
 400 storm. Then we determine the center of the storm by averaging the locations of these two  
 401 local extrema. Empirically, we set  $\sigma = 2$  for the Gaussian kernel and  $\sigma_1 = 2$ ,  $\sigma_2 = 8$  for  
 402 the Laplacian kernel, for the best performance. In most cases, the positions of maximum  
 403 vorticity closely align with those of minimum sea level pressure. However, when a TC’s  
 404 symmetric structure is not well maintained, this approach aids in stabilizing the outcomes  
 405 and provides robust tracking. It should be noted that Pangu-Weather includes an algorithm  
 406 for tracking TCs, which is based on relative vorticity, geopotential thickness, and 10-m wind  
 407 speed. Our tracking method delivers comparable results, while requiring fewer data inputs.  
 408 This advantage makes our approach generalizable to other weather forecasting systems that  
 409 may lack specific variables (for example, ForecastNet does not output vorticity; see (Pathak  
 410 et al., 2022)).

### 411 3.5 Bias correction of TC maximum wind speed

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. This leads to an underestimation of the maximum wind speeds of the TC. To address this problem, we develop a separate machine learning model to correct this bias based on the characteristics of the storm and the environment within the inner region of the storm. Due to the recurrent nature of this problem, where the wind speed at time  $t$  is highly correlated with the wind speed at previous time steps, we formulate the problem using a recurrent neural network based on long-short-term memory (LSTM). The structure of our network is shown in Figure 5 subplot b, where the bias correction stage is formulated as:

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

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

412 In Eq. 3, we use the term "raw" to indicate the intensity directly simulated from TC-  
 413 GEN, which is expected to have the same statistics as those in the ERA5 reanalysis data.  
 414 The term "true" is used to denote the real intensity, which has statistics similar to IBTrACs.  
 415 Thus,  $\mathbf{I}_{\text{raw}}$  and  $\mathbf{I}_{\text{true}}$  represent the maximum wind speed before and after bias correction.  
 416  $\mathbf{E}(\mathbf{x}(t))$  represents the environmental predictors extracted from the area surrounding the  
 417 TC center, which is located at  $\mathbf{x}(t)$  at time  $t$ .  $\mathbf{F}$  represents a machine learning model with  
 418 learnable weights  $\theta$ .

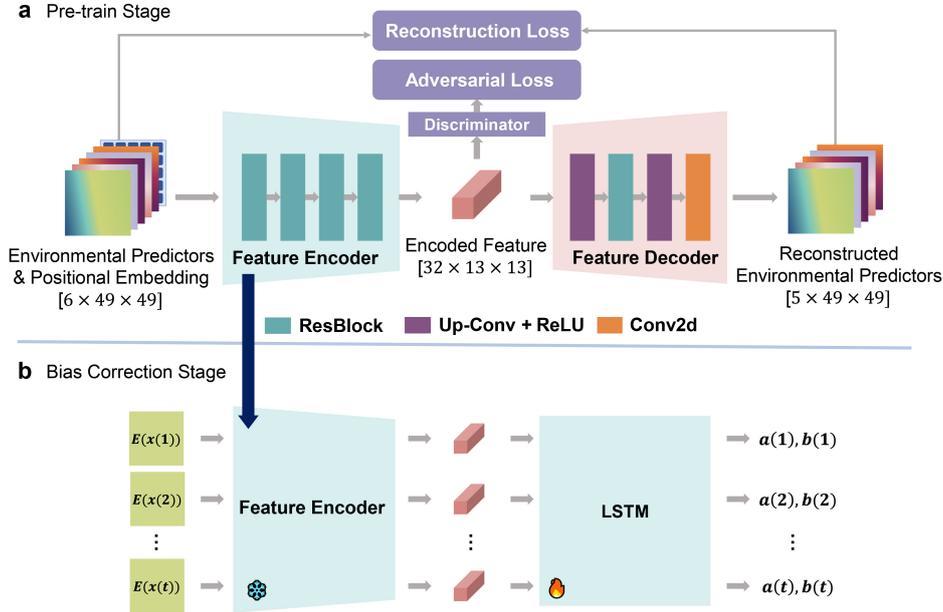
419 We use five environmental variables (mean sea level pressure, u- and v- component of  
 420 10m wind speed, relative humidity at 850 hpa and temperature at 850 hpa) that have been  
 421 identified as the most important predictors of the intensity of TC (Jing & Lin, 2019). At  
 422 each time step, the five variables within a circular region, covering a radius equivalent to  
 423 49 grid cells and centered at the TC, are combined with a positional embedding (Lam et  
 424 al., 2022) to create a raw input. We then convert the raw input to a feature vector using  
 425 a feature encoder comprised of 4 ResBlocks, and feed these feature vectors into an LSTM,  
 426 predicting  $a(t), b(t)$  at each time step.

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

433 To bridge the domain gap, we pre-train the feature encoder so that the extracted feature  
 434 is informative to reconstruct the raw ERA5 environment, yet indistinguishable in terms of  
 435 its source, i.e. whether it comes from ERA5 or Pangu-Weather. Such properties are achieved  
 436 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  
 438 by a feature decoder to reconstruct the original features using an L1 loss. Additionally,  
 439 we introduce an adversarial loss that performs a binary classification task, attempting to  
 440 discern the source of the feature (ERA5 or Pangu-Weather). The feature encoder is trained  
 441 to deceive the discriminator to the extent that it cannot identify the source of the feature.  
 442 As a result, the extracted features become domain-agnostic after this stage and the domain  
 443 gap is mitigated.

444 Overall, we start by training the auto-encoder architecture using a combination of  
 445 environment maps sampled from both the ERA5 and Pangu-Weather output (Figure 5  
 446 subplot a). Once the autoencoder converges, we discard the decoder, freeze the encoder  
 447 weights for feature extraction, and only update the LSTM weights (Figure 5 subplot b).  
 448 To train the LSTM, we divide the historical data of the TCs from 1979 to 2021, randomly  
 449 selecting 80% of the TCs for training and reserving the remaining 20% for testing. We use  
 450 the real maximum wind speed data from IBTrACS as ground truth. The LSTM is trained  
 451 with an AdamW optimizer (Loshchilov & Hutter, 2017), with a Huber loss and a learning  
 452 rate of 0.0001 over 10 iterations.

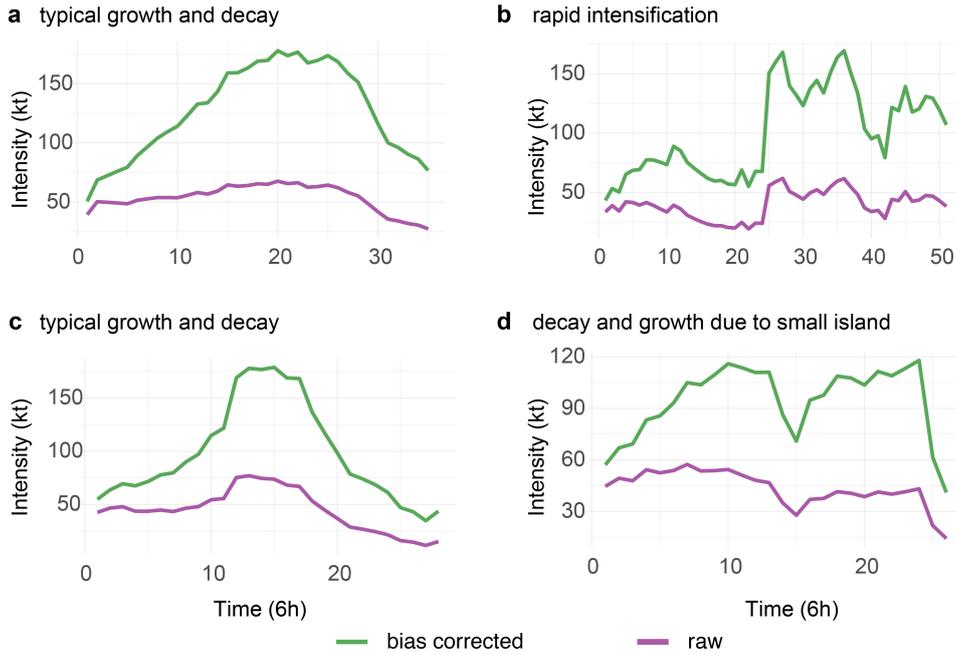


**Figure 5. Structure of the intensity bias correction model** The model consists of two stages, (a) a pre-trained stage to bridge domain gap between ERA5 and Pangu-Weather output so that all features become domain-agnostic; and (b) a bias correction stage that adjust the maximum wind speed using a Long Short-Term Memory (LSTM) model.

453 The performance of the LSTM-based bias correction model is illustrated in Figure 6,  
 454 which shows four cases that represent various characteristics of the storm. These include  
 455 typical growth and decay, rapid intensification, and storms that weaken after hitting an  
 456 island but subsequently strengthen after moving over the ocean.

### 457 3.6 Extracting the radius of outer size from each storm

458 The destructive potential of a TC is related to both its maximum sustained wind speed and  
 459 the radial extent of the near-surface wind, the latter typically measured by the outer



**Figure 6. Examples of bias correction in 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.

460 size of the TC (Powell & Reinhold, 2007; Irish et al., 2008). Therefore, it is important  
 461 to assess how well TC-GEN can replicate the realistic horizontal wind structures of TCs.  
 462 Furthermore, since TC seeds are artificially integrated into the surrounding environment,  
 463 successfully reproducing the historical distribution of the outer size of TCs can, to some  
 464 extent, prove that integration based on weighted Poisson blending is feasible.

465 Previous studies have shown that reanalysis datasets, including the ERA5 reanalysis  
 466 used in this study, are capable of effectively capturing the TC outer size, which is defined  
 467 by radii where convection is minimal and the atmosphere maintains a radiative–subsidence  
 468 balance (Schenkel et al., 2017). In earlier studies, multiple thresholds for azimuthal mean  
 469 tangential winds have been used, such as 2 m/s, 6 m/s, 8 m/s, 9 m/s, and 12 m/s (Bian et  
 470 al., 2021). Here, we use three size metrics based on the radii where the wind diminishes to  
 471 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  
 472 both the smallest and the largest metrics.

473 We identify the outer size of the TC following the procedures outlined in (Schenkel et  
 474 al., 2017). The procedure involves the following five main steps: 1) remove the background  
 475 flow from the surface wind field, which is empirically defined as the storm translation vector  
 476 rotated 20 deg cyclonically and diminished by a factor of 0.55, as described in the method  
 477 by (Lin et al., 2012); 2) establish a TC-centered polar coordinate system, where the cyclone  
 478 center serves as the origin, and the grid spacing in the radial direction is set at 0.5 times  
 479 the archived reanalysis grid spacing; 3) compute the tangential wind at the grid points of  
 480 the reanalysis data for each annular region of the polar coordinate system; 4) remove all  
 481 radial points if the data asymmetry parameter exceeds 0.5, following (Chavas & Vigh, 2014);  
 482 5) calculate the gridded radial profile of azimuthal-mean 10-m tangential wind using cubic  
 483 Hermite polynomial interpolation, based on which we then identify the radii of at which the

484 10-meter azimuthal-mean tangential wind speeds decrease to a certain threshold (12 m/s, 6  
485 m/s, and 2 m/s).

## 486 **4 TC-GEN Evaluation**

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

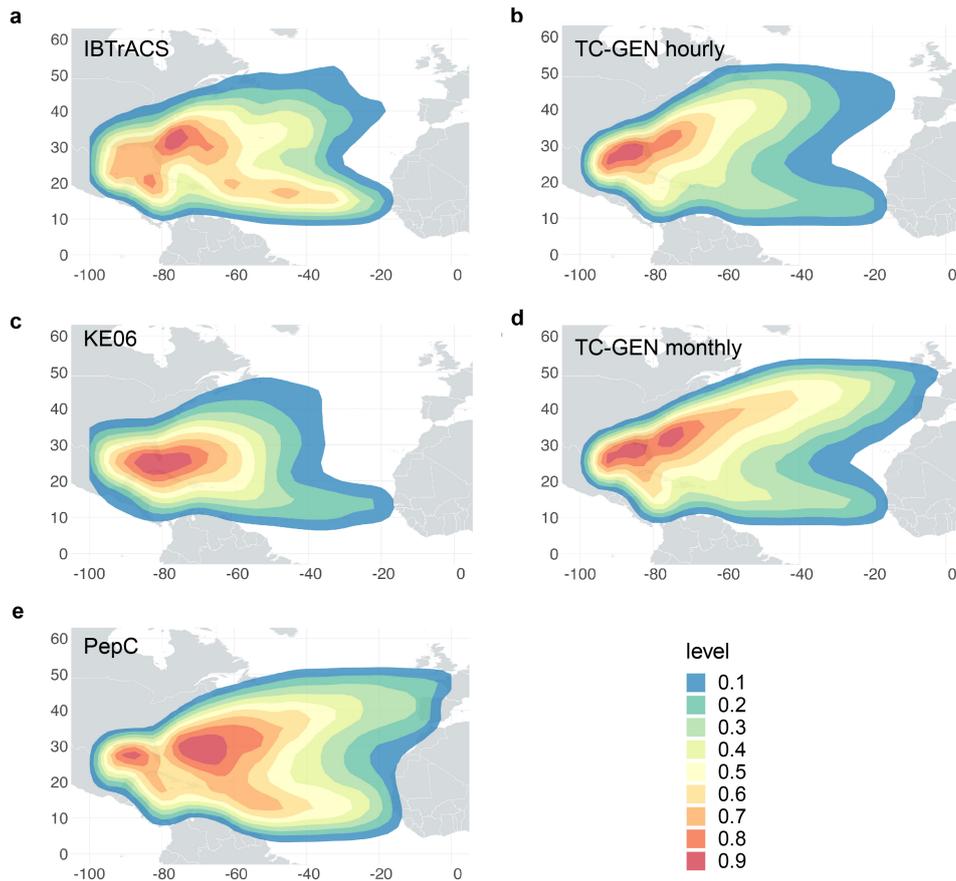
### 501 **4.1 Track Density**

502 Figure 7 compares the simulated tracks that are initiated with both hourly and monthly  
503 data, with observed tracks and simulated tracks using KE08 and PepC. The colors represent  
504 the spatial track density normalized to the maximum of the basin. We show that both sets of  
505 simulated tracks replicate the typical recurving pattern seen in the observed tracks relatively  
506 well, which is comparable to KE08 and PepC. Compared to observations, TC-GEN simulated  
507 tracks exhibit a negative bias in the main development region, mainly stemming from the  
508 negative bias in the genesis component of PepC within this region. We test this hypothesis  
509 with a sensitivity test by resampling the genesis according to the spatial distribution of  
510 historical genesis locations. After sampling, we find that the simulated tracks effectively  
511 capture the hotspots in the main development region, the southeast US coast, and the Gulf  
512 of Mexico, although there is a slight positive bias in the Gulf of Mexico with hourly data  
513 and in the West Caribbean Sea with monthly data (Figure S1).

514 We further evaluate the performance of TC-GEN by comparing the 6-hourly north-  
515 south and east-west displacements of simulated and observed tracks, which serves as a  
516 means to assess TC-GEN's performance in simulating individual tracks. The results are  
517 shown in Figure 8. All simulated data sets are largely consistent with the observations.  
518 For simulated tracks initiated with hourly data, there is a slight positive bias for positive  
519 meridional displacement, a slight negative bias for negative meridional displacement; and  
520 correspondingly, a negative bias for positive zonal displacements and a positive bias for  
521 negative zonal displacements, which could come from fewer recurvations in simulated storm  
522 tracks that originate in the main development region. Additionally, positive biases in merid-  
523 ional displacements may also arise from the deviations in eastward moving tracks at high  
524 latitude (near Europe), where they should have been terminated because of their low wind  
525 intensity. These patterns are also seen in simulated tracks that are initiated with monthly  
526 reanalysis data. In general, both datasets exhibit comparable performance to existing down-  
527 scaling methods, with simulated tracks initiated using hourly data showing slightly better  
528 performance than those initiated with monthly data.

### 529 **4.2 Landfall Frequency**

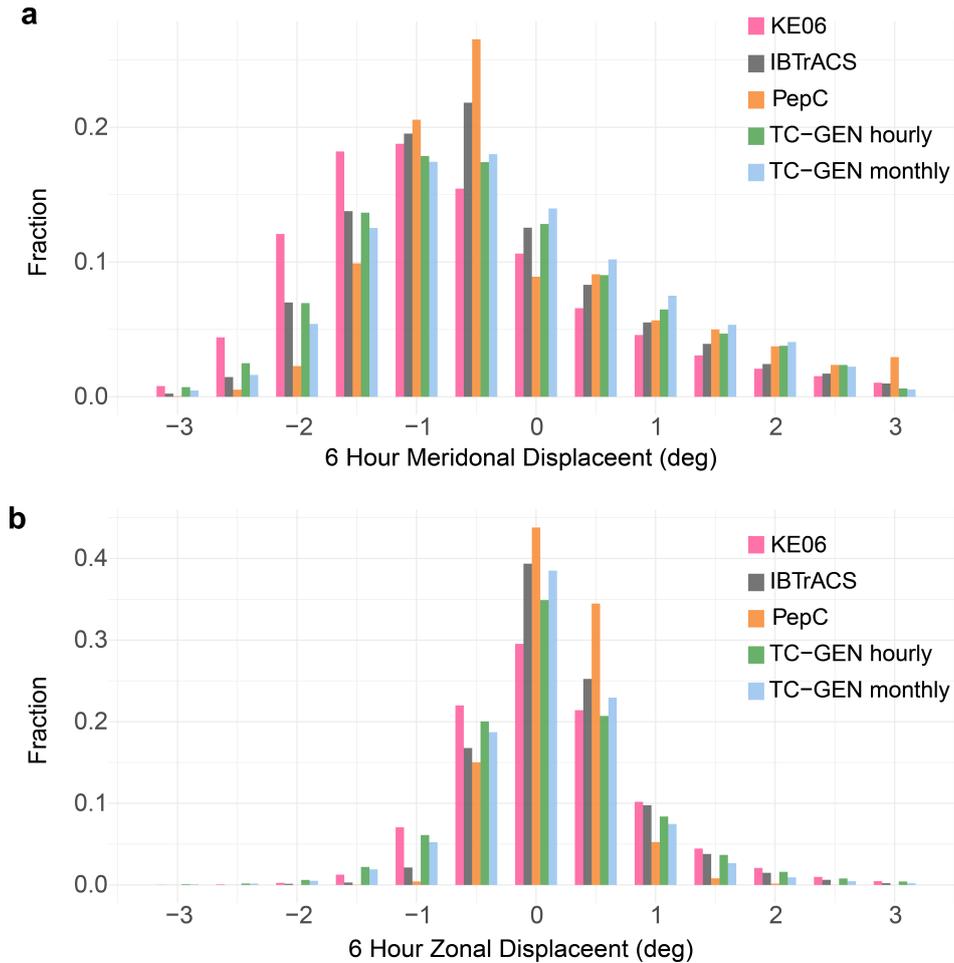
530 We examine the annual regional landfall frequency at coastal locations along the North  
531 Atlantic coastline. To help indicate locations, a total of 186 mileposts (MPs) are defined



**Figure 7. Track density** Track density is calculated as the accumulated number of TC passes into each  $0.75 \times 0.75$  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.

532 following Vickery et al. (2000) (Vickery et al., 2000), as shown in Figure 9, to cover the  
 533 coastline with 100-km spacing along the Mexican coastline and 50-km spacing along the US  
 534 coastline. Landfall is defined as simulated or observed storms that approach within 50 km  
 535 of each coastal milepost. The results presented in Figure 9b are based on simulated tracks  
 536 initiated using both hourly and monthly data. Furthermore, results from KE08- and PepC-  
 537 simulated tracks are included for comparison. As a reference, the historical annual landfall  
 538 rate between 1900 and 2022 is shown with shading that indicates the associated sampling  
 539 error at each milepost. The sampling error for the annual rate of each gate is determined  
 540 by calculating the total number and standard error of storms that cross each gate over the  
 541 entire record and then dividing both the mean and the error bars by the number of years.

542 Due to the different annual total frequencies in the simulated and observed track  
 543 datasets, which can significantly influence landfall frequency, we adjust the annual rate  
 544 to a uniform 13 storms per year in all datasets to form a fair comparison. Our results  
 545 indicate that the simulated tracks, from daily and monthly data, can reproduce the overall  
 546 pattern in the observations, showing a performance comparable to that of KE08 and PepC.  
 547 The annual landfall rates simulated by TC-GEN exhibit correlations with observations of  
 548 0.79 and 0.82 for the tracks initiated with hourly and monthly data, respectively, which are

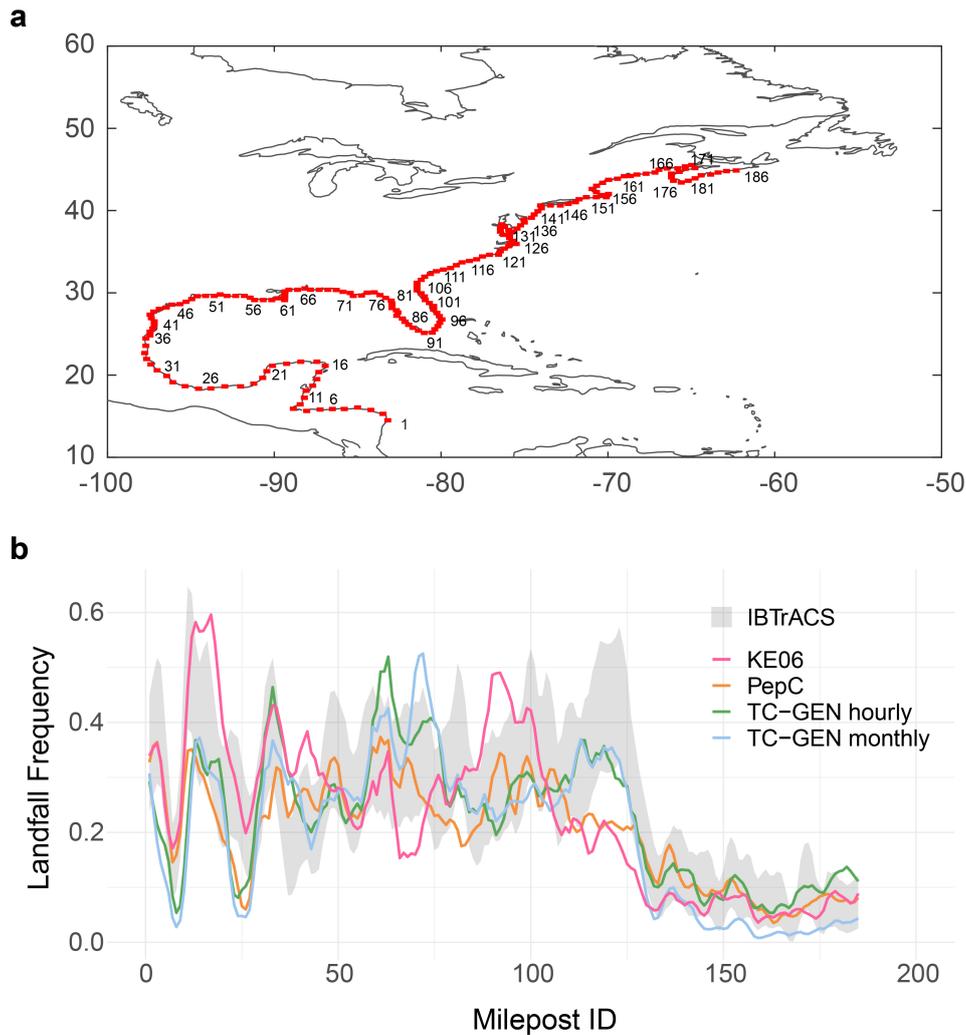


**Figure 8. 6 hour zonal and meridional displacement** Comparison of probability density functions of 6 hour (a) meridional and (b) zonal displacements between simulated tracks and observations. Simulated tracks initiated with hourly data and monthly data are both presented and compared with KE08 and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

549 comparable to 0.83 and 0.79 for PepC and KE08, respectively. Both TC-GEN simulated  
 550 data sets capture the landfall frequency at MP 100-125, which is overlooked by existing  
 551 methods, while they underestimate the landfall frequency below MP 25, possibly due to  
 552 fewer genesis in the main development region.

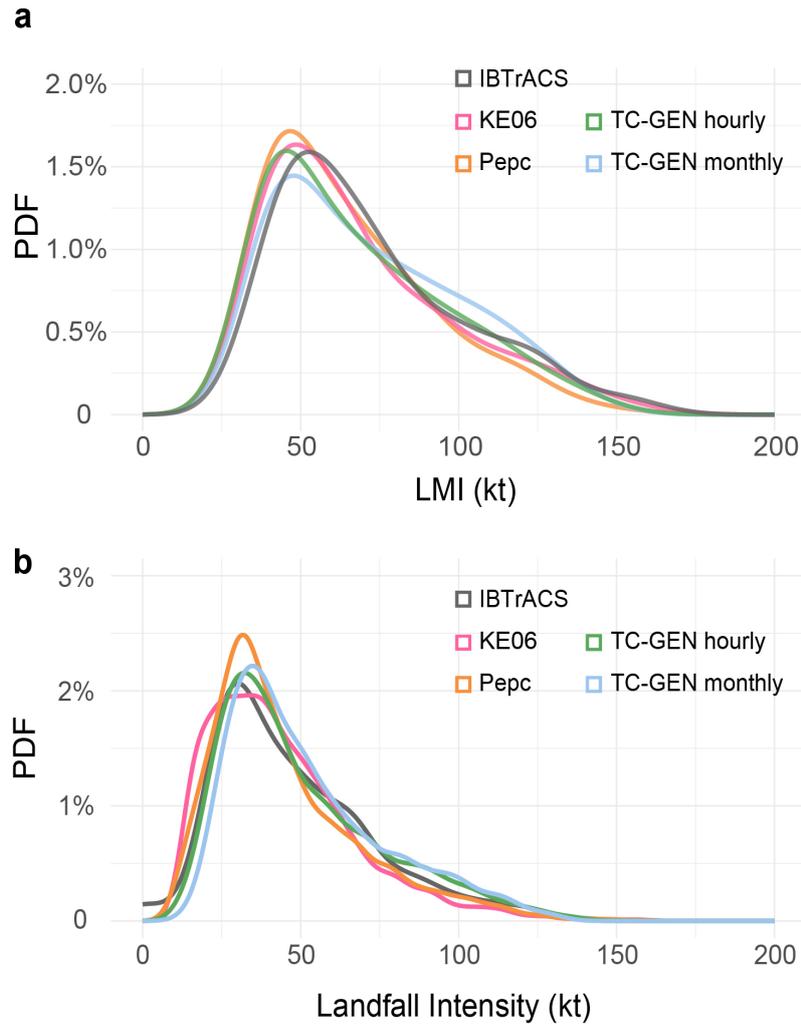
### 553 4.3 Lifetime Maximum Intensity and Landfall Intensity

554 We analyze the simulated TC intensity using two metrics, the lifetime maximum in-  
 555 tensity and landfall intensity. The LMI distribution serves as a representation of the TC  
 556 intensity climatology. A successful simulation should be able to reproduce the bimodal  
 557 distribution with a shoulder feature around 120 kt, which is associated with storms that  
 558 undergo rapid intensification (C.-Y. Lee et al., 2016). We show in Figure 10 that after bias  
 559 correction, both sets of TC-GEN are capable of reproducing a realistic distribution of the  
 560 observed LMI. TC-GEN tracks initiated with hourly data better simulate the tail of the LMI  
 561 distribution for LMI greater than 75 kt; however, there is a negative bias in LMI for storms  
 562 with LMI less than 75 kt, which are mostly moderate storms that do not undergo rapid



**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.

563 intensification. TC-GEN tracks initiated with monthly data overestimate the proportion of  
 564 storms with an LMI greater than 75 kt. The biases might originate from the intensity bias  
 565 correction model. In order to have a reasonable fraction of storms undergo rapid intensi-  
 566 fication, the bias correction model prioritizes strong storms, which could lead to a higher  
 567 proportion of storms becoming more intense than they should be. Similar patterns are also  
 568 seen in the distribution of the landfall intensity. Both data sets initiated with hourly and  
 569 monthly data exhibit a positive bias in landfall intensity greater than 75 knots. This bias is  
 570 also likely to be attributed to the intensity correction model, which tends to overestimate  
 571 the intensity of the storm when the storm has already weakened.



**Figure 10. Lifetime maximum intensity 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.

572

#### 4.4 Outer Size

573

574

575

576

577

578

579

580

581

582

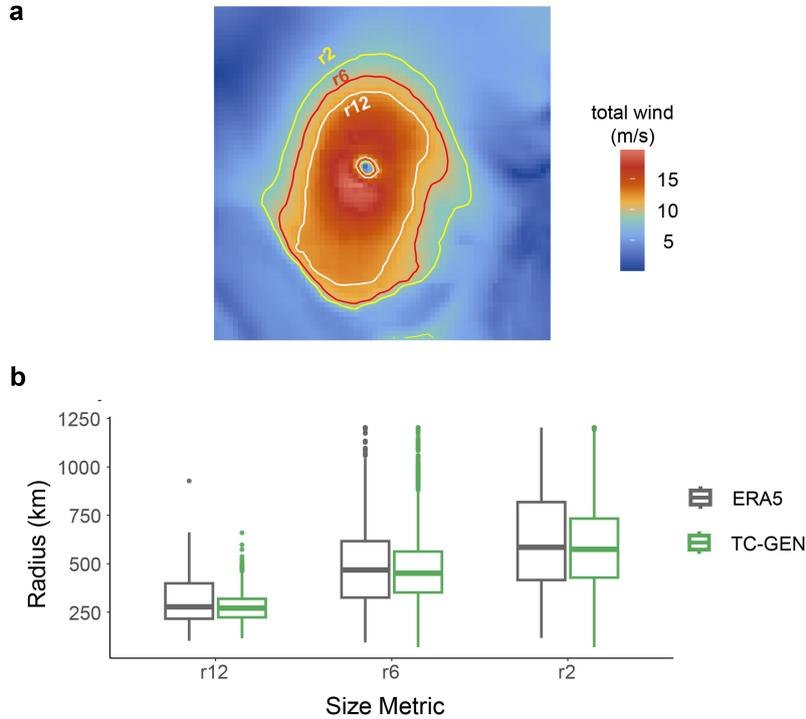
583

584

585

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, and ERA5 shows an improved representation of the outer size compared to previous versions of ERA (Bian et al., 2021). Here, we examine how the outer size of the TCs compares between TC-GEN simulated tracks and historical storms, where outer sizes of historical TCs are derived from the corresponding ERA5 reanalysis data. As the statistical downscaling approach does not provide this output, we do not have data from 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 outer size, with a discrepancy of approximately 10 km, which is even lower than the uncertainties in outer size arising from different reanalysis datasets (Bian et al., 2021). However, the

586 standard deviations for all three size metrics in TC-GEN-simulated TCs are smaller than  
 587 those observed in historical storms, particularly for r12, which represents the largest metric  
 588 that defines the outer size of storms. This smaller variation may arise from the bias in the  
 589 horizontal wind structures of TCs at genesis, which requires further investigation.



**Figure 11. 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-meter 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 ERA reanalysis. 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.

## 590 5 Discussion

### 591 5.1 Sensitivity to reanalysis data

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

598 Furthermore, we evaluate the suitability of applying TC-GEN to reanalysis data sets  
 599 other than ERA5 reanalysis. Using NCEP GFS (see 2.3, we find that the results are unsat-  
 600 isfactory (as illustrated in Figure S2), which is likely attributed to the well-known domain  
 601 gap problem in deep learning. In a deep learning model, typically the first several layers  
 602 of the neural network are responsible for extracting features from raw data. These layers  
 603 are trained to identify relevant features that help in downstream tasks, such as recogniz-

ing distinctive patterns, reducing noise, and achieving specific invariances. 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. Based on this discussion, an interesting future direction involves devising a generalized ML-based weather forecasting system that performs reasonably well on various reanalysis datasets. This could be accomplished, for example, by training the model on multiple reanalysis datasets 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. 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 200 km) to 89 grid cells (approximately 1200 km). Our sensitivity tests reveal that the results are relatively robust when the radius falls within the range of 49 to 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. In the development of this model, we also test 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. One limitation of the current bias correction model is that it tends to overestimate TC maximum intensity when the storm is decaying, which

partially explains the positive bias in the landfall intensity distribution. This bias 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). Consequently, the model tends to generate more storms that grow at a higher rate while decaying at a slower rate. Future work should focus on improving the bias correction model to effectively handle storms that undergo rapid intensification and those that do not, with the aim of achieving optimal performance in terms of both lifetime maximum intensity and landfall intensity.

## 5.5 Asymmetric TC wind field

Horizontal asymmetric TC wind fields are important for disaster management and regional risk assessment. In addition to the maximum wind speed, accurate estimation of the TC wind field is essential to identify at-risk populations and assess potential climate-related threats. One notable advantage of TC-GEN is its ability to directly output spatial asymmetric characteristics of the TC surface wind field. Although raw simulated winds are often underestimated as TC structures cannot be fully resolved at a resolution of  $0.25^\circ \times 0.25^\circ$ , there are still several ways to adjust the intensity of the wind and effectively use the simulated asymmetries. For example, with the adjusted maximum wind intensity and the extracted outer size of the TC, a parametric wind model can be applied, such as the model developed in Chavas et al. (2015) (Chavas et al., 2015), to generate the complete asymmetric wind profile of 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 a novel machine learning-based approach to the synthetic downscaling of tropical cyclones. This approach, which we refer to as "TC-GEN", leverages the recent advances in machine learning-based global weather models. The machine learning-based high-resolution global weather model (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 four key steps: 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, and the correction maximum wind speed biases using a long-short-term memory model. 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.

The recently introduced ML-GWM has significant potential for improvement with ongoing advances in machine learning and artificial intelligence, aiming for 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 tropical cyclone risk assessment, we believe that machine learning-based global weather models will play a crucial role in future climate studies.

## Open Research Section

Pangu-Weather trained models can be downloaded from the public GitHub repository at <https://github.com/198808xc/Pangu-Weather> (<https://doi.org/10.5281/zenodo.7678849>). Tropical cyclone observations are obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) project at <https://www.ncei.noaa.gov/products/international-best-track-archive>. Historical ERA5 reanalysis data (both monthly and 6-hourly) are obtained from the ECMWF climate data store. Historical NCEP reanalysis data is downloaded at <https://rda.ucar.edu/datasets/ds084.1/>.

## Acknowledgments

K. Emanuel's research is part of the MIT Climate Grand Challenge on Weather and Climate Extremes. This research received support by the generosity of Eric and Wendy Schmidt by recommendation of Schmidt Futures as part of its Virtual Earth System Research Institute (VESRI).

## References

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4), 433–459.
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2022). Pangu-weather: A 3d high-resolution model for fast and accurate global weather forecast. *arXiv preprint arXiv:2211.02556*.
- 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.
- 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.
- 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.
- 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.
- Chavas, D. R., & Vigh, J. (2014). Qscat-r: The quikscat tropical cyclone radial structure dataset. *NCAR Tech. Note TN-5131STR*.
- 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. (2017). Assessing the present and future probability of hurricane harvey's rainfall. *Proceedings of the National Academy of Sciences*, 114(48), 12681–12684.
- Emanuel, K., DesAutels, C., Holloway, C., & Korty, R. (2004). Environmental control of tropical cyclone intensity. *Journal of the atmospheric sciences*, 61(7), 843–858.
- 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.
- Emanuel, K., Sundararajan, R., & Williams, J. (2008). Hurricanes and global warming:

- 752 Results from downscaling ipcc ar4 simulations. *Bulletin of the American Meteorological*  
753 *Society*, 89(3), 347–368.
- 754 Emanuel, K. A. (1988). The maximum intensity of hurricanes. *J. Atmos. Sci*, 45(7),  
755 1143–1155.
- 756 for Environmental Prediction/National Weather Service/NOAA/US Department of Com-  
757 merce, N. C. (2015). Ncep gfs 0.25 degree global forecast grids historical archive.  
758 *Research Data Archive at the National Center for Atmospheric Research, Computa-*  
759 *tional and Information Systems Laboratory.*
- 760 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., . . . oth-  
761 ers (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological*  
762 *Society*, 146(730), 1999–2049.
- 763 Huang, M., Wang, Q., Jing, R., Lou, W., Hong, Y., & Wang, L. (2022). Tropical cyclone  
764 full track simulation in the western north pacific based on random forests. *Journal of*  
765 *Wind Engineering and Industrial Aerodynamics*, 228, 105119.
- 766 Huang, M., Wang, Q., Li, Q., Jing, R., Lin, N., & Wang, L. (2021). Typhoon wind hazard  
767 estimation by full-track simulation with various wind intensity models. *Journal of*  
768 *Wind Engineering and Industrial Aerodynamics*, 218, 104792.
- 769 Huang, M., Wang, Q., Liu, M., Lin, N., Wang, Y., Jing, R., . . . Lou, W. (2022). Increasing  
770 typhoon impact and economic losses due to anthropogenic warming in southeast china.  
771 *Scientific reports*, 12(1), 14048.
- 772 Irish, J. L., Resio, D. T., & Ratchiff, J. J. (2008). The influence of storm size on hurricane  
773 surge. *Journal of Physical Oceanography*, 38(9), 2003–2013.
- 774 James, M., & Mason, L. (2005). Synthetic tropical cyclone database. *Journal of waterway,*  
775 *port, coastal, and ocean engineering*, 131(4), 181–192.
- 776 Jing, R., & Lin, N. (2019). Tropical cyclone intensity evolution modeled as a dependent  
777 hidden markov process. *Journal of Climate*, 32(22), 7837–7855.
- 778 Jing, R., & Lin, N. (2020). An environment-dependent probabilistic tropical cyclone model.  
779 *Journal of Advances in Modeling Earth Systems*, 12(3), e2019MS001975.
- 780 Jing, R., Lin, N., Emanuel, K., Vecchi, G., & Knutson, T. R. (2021). A comparison  
781 of tropical cyclone projections in a high-resolution global climate model and from  
782 downscaling by statistical and statistical-deterministic methods. *Journal of Climate*,  
783 34(23), 9349–9364.
- 784 Kareem, A., Hu, L., Guo, Y., & Kwon, D.-K. (2019). Generalized wind loading chain: Time-  
785 frequency modeling framework for nonstationary wind effects on structures. *Journal*  
786 *of Structural Engineering*, 145(10), 04019092.
- 787 Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., & Schreck, C. J. (2018). Interna-  
788 tional best track archive for climate stewardship (ibtracs) project, version 4.v04r00.  
789 doi: doi:10.25921/82ty-9e16
- 790 Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010).  
791 The international best track archive for climate stewardship (ibtracs) unifying tropical  
792 cyclone data. *Bulletin of the American Meteorological Society*, 91(3), 363–376.
- 793 Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., . . .  
794 others (2022). Graphcast: Learning skillful medium-range global weather forecasting.  
795 *arXiv preprint arXiv:2212.12794*.
- 796 Lee, C.-Y., Camargo, S. J., Sobel, A. H., & Tippett, M. K. (2020). Statistical–dynamical  
797 downscaling projections of tropical cyclone activity in a warming climate: Two di-  
798 verging genesis scenarios. *Journal of Climate*, 33(11), 4815–4834.
- 799 Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2016). Rapid intensification  
800 and the bimodal distribution of tropical cyclone intensity. *Nature communications*,  
801 7(1), 10625.
- 802 Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2018). An environmentally  
803 forced tropical cyclone hazard model. *Journal of Advances in Modeling Earth Systems*,  
804 10(1), 223–241.
- 805 Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., . . . others (2023).  
806 Climate change 2023: synthesis report. contribution of working groups i, ii and iii to

- 807 the sixth assessment report of the intergovernmental panel on climate change.
- 808 Lin, N., & Chavas, D. (2012). On hurricane parametric wind and applications in storm  
809 surge modeling. *Journal of Geophysical Research: Atmospheres*, *117*(D9).
- 810 Lin, N., Emanuel, K., Oppenheimer, M., & Vanmarcke, E. (2012). Physically based assess-  
811 ment of hurricane surge threat under climate change. *Nature Climate Change*, *2*(6),  
812 462–467.
- 813 Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint*  
814 *arXiv:1711.05101*.
- 815 Marsooli, R., Lin, N., Emanuel, K., & Feng, K. (2019). Climate change exacerbates hurricane  
816 flood hazards along us atlantic and gulf coasts in spatially varying patterns. *Nature*  
817 *communications*, *10*(1), 3785.
- 818 Meiler, S., Vogt, T., Bloemendaal, N., Ciullo, A., Lee, C.-Y., Camargo, S. J., ... Bresch,  
819 D. N. (2022). Intercomparison of regional loss estimates from global synthetic tropical  
820 cyclone models. *Nature Communications*, *13*(1), 6156.
- 821 Mendelsohn, R., Emanuel, K., Chonabayashi, S., & Bakkensen, L. (2012). The impact  
822 of climate change on global tropical cyclone damage. *Nature climate change*, *2*(3),  
823 205–209.
- 824 Nam, H., Lee, H., Park, J., Yoon, W., & Yoo, D. (2021). Reducing domain gap by reducing  
825 style bias. In *Proceedings of the ieee/cvf conference on computer vision and pattern*  
826 *recognition* (pp. 8690–8699).
- 827 Nandi, A., Mandal, A., Wilson, M., & Smith, D. (2016). Flood hazard mapping in jamaica  
828 using principal component analysis and logistic regression. *Environmental Earth Sci-*  
829 *ences*, *75*, 1–16.
- 830 Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., ...  
831 Anandkumar, A. (2022). Fourcastnet: A global data-driven high-resolution weather  
832 model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*.
- 833 Pérez, P., Gangnet, M., & Blake, A. (2003). Poisson image editing. *ACM SIGGRAPH 2003*  
834 *Papers*.
- 835 Pörtner, H.-O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., ... others (2022).  
836 Climate change 2022: Impacts, adaptation and vulnerability. *IPCC Sixth Assessment*  
837 *Report*.
- 838 Powell, M. D., & Reinhold, T. A. (2007). Tropical cyclone destructive potential by integrated  
839 kinetic energy. *Bulletin of the American Meteorological Society*, *88*(4), 513–526.
- 840 Ruan, Z., & Wu, Q. (2022). Relationship between size and intensity in north atlantic  
841 tropical cyclones with steady radii of maximum wind. *Geophysical Research Letters*,  
842 *49*(3), e2021GL095632.
- 843 Schenkel, B. A., Lin, N., Chavas, D., Oppenheimer, M., & Brammer, A. (2017). Evaluating  
844 outer tropical cyclone size in reanalysis datasets using quikscat data. *Journal of*  
845 *Climate*, *30*(21), 8745–8762.
- 846 Tippett, M. K., Camargo, S. J., & Sobel, A. H. (2011). A poisson regression index for  
847 tropical cyclone genesis and the role of large-scale vorticity in genesis. *Journal of*  
848 *Climate*, *24*(9), 2335–2357.
- 849 Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., ... Birchfield,  
850 S. (2018). Training deep networks with synthetic data: Bridging the reality gap by  
851 domain randomization. In *Proceedings of the ieee conference on computer vision and*  
852 *pattern recognition workshops* (pp. 969–977).
- 853 Vickery, P., Skerlj, P., & Twisdale, L. (2000). Simulation of hurricane risk in the us using  
854 empirical track model. *Journal of structural engineering*, *126*(10), 1222–1237.
- 855 Wei, L., Zhang, S., Gao, W., & Tian, Q. (2018). Person transfer gan to bridge domain gap  
856 for person re-identification. In *Proceedings of the ieee conference on computer vision*  
857 *and pattern recognition* (pp. 79–88).

# TC-GEN: Data-driven Tropical Cyclone Downscaling using Machine Learning-Based High-resolution Weather Model

Renzhi Jing<sup>1,2</sup>, Jianxiong Gao<sup>3</sup>, Yunuo Cai<sup>4</sup>, Dazhi Xi<sup>5</sup>, Yinda Zhang<sup>6</sup>,  
Yanwei Fu<sup>4</sup>, Kerry Emanuel<sup>7</sup>, Noah S. Diffenbaugh<sup>1,8,9</sup>, Eran Bendavid<sup>2,10</sup>

<sup>1</sup>Woods Institute for the Environment, Stanford University, Stanford, CA, USA

<sup>2</sup>School of Medicine, Stanford University, Stanford, CA, USA

<sup>3</sup>Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai, China

<sup>4</sup>School of Data Science and MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China

<sup>5</sup>Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA

<sup>6</sup>Google LLC, 1600 Amphitheatre Pkwy, Mountain View, CA, USA

<sup>7</sup>Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology,

Cambridge, MA, USA

<sup>8</sup>Department of Earth System Science, Stanford University, Stanford, CA, USA

<sup>9</sup>Doerr School of Sustainability, Stanford University, Stanford, CA, USA

<sup>10</sup>Department of Health Policy, Stanford University, Stanford, CA, USA

## Key Points:

- We develop a novel approach for tropical cyclone downscaling using a machine learning-based high-resolution global weather model.
- We generate and integrate synthetic tropical cyclone seeds into the surrounding environment using a data-driven approach.
- The new tropical cyclone downscaling approach is capable of simulating the two-way interactions between storms and the environment as a unified system.

---

Corresponding author: Renzhi Jing, [rjing@stanford.edu](mailto:rjing@stanford.edu)

Corresponding author: Yanwei Fu, [yanweifu@fudan.edu.cn](mailto:yanweifu@fudan.edu.cn)

## 24 Abstract

25 Synthetic downscaling of tropical cyclones (TCs) is critically important to estimate the  
 26 long-term hazard of rare high-impact storm events. Existing downscaling approaches rely  
 27 on statistical or statistical-deterministic models that are capable of generating large samples  
 28 of synthetic storms with characteristics similar to observed storms. However, these models  
 29 do not capture the complex two-way interactions between a storm and its environment. In  
 30 addition, these approaches either necessitate a separate TC size model to simulate storm  
 31 size or involve post-processing to introduce asymmetries in the simulated surface wind. In  
 32 this study, we present an innovative data-driven approach for TC synthetic downscaling.  
 33 Using a machine learning-based high-resolution global weather model (ML-GWM), our ap-  
 34 proach is able to simulate the full life cycle of a storm with asymmetric surface wind that  
 35 accounts for the two-way interactions between the storm and its environment. This ap-  
 36 proach consists of multiple components: a data-driven model for generating synthetic TC  
 37 seeds, a blending method that seamlessly integrate storm seeds into the surrounding while  
 38 maintain the seed structure, and a recurrent neural network-based model for correcting the  
 39 biases in maximum wind speed. Compared to observations and synthetic storms simulated  
 40 using existing statistical-deterministic and statistical downscaling approaches, our method  
 41 shows the ability to effectively capture many aspects of TC statistics, including track den-  
 42 sity, landfall frequency, landfall intensity, and outermost wind extent. Taking advantage of  
 43 the computational efficiency of ML-GWM, our approach shows substantial potential for TC  
 44 regional hazard and risk assessment.

## 45 Plain Language Summary

46 Tropical cyclones (TCs) cause significant destruction each year. It is crucial to accu-  
 47 rately assess the risks they present, but this is challenging due to a scarcity of historical  
 48 data. A commonly used approach involves creating a large number of synthetic TCs that  
 49 share key characteristics with real storms, enabling an effective regional risk assessment.  
 50 However, traditional synthetic TC generation approaches do not capture the complex inter-  
 51 actions between storms and their larger-scale environment. Furthermore, these approaches  
 52 do not adequately represent the asymmetric structure of TCs, despite the crucial role that  
 53 they play in modeling storm-related hazards such as rainfall and surges. Recently, advances  
 54 in machine learning-based global weather forecasting (ML-GWM) have provided highly ac-  
 55 curate and efficient high-resolution global weather forecasts that surpass the capabilities of  
 56 conventional numerical weather forecasting. In this study, we introduce a novel synthetic TC  
 57 generation approach, which we call the synthetic TC-GENerative Model (or "TC-GEN"),  
 58 leveraging the state-of-the-art ML-GWM. We show that TC-GEN can generate a large num-  
 59 ber of synthetic storms that allow the interaction between the storm and its environment.  
 60 We evaluate the performance of TC-GEN in various aspects, including several landfall char-  
 61 acteristics, which are of the most importance for local TC risk analysis. Our study also  
 62 serves as a compelling example of the transformative impact of machine learning and data  
 63 science in revolutionizing climate studies during the era of artificial intelligence.

## 64 1 Introduction

65 Tropical cyclones (TCs) are among the most destructive natural disasters, causing  
 66 substantial damage and losses in multiple ocean basins annually. In a warming climate, it is  
 67 projected that TCs are likely to become more intense, with an expected increase in both the  
 68 peak maximum wind speed and the proportion of strong TCs in the future (Pörtner et al.,  
 69 2022; H. Lee et al., 2023). Accurate assessment of TC tracks and intensities is fundamental to  
 70 reducing the impacts of landfalling storms. However, with around 90 storms occurring every  
 71 year and an average of only 20 making landfall, this task is challenging due to the shortage of  
 72 historical data required for regional risk assessment. To overcome data deficiency, a widely  
 73 used approach is to generate synthetic TCs that are capable of responding to various climate

74 conditions (K. Emanuel et al., 2008; C.-Y. Lee et al., 2018; Jing & Lin, 2020). Using large  
75 samples of synthetic storms, including extreme events with extended return periods, enables  
76 a comprehensive risk assessment for specific regions in both current and future climates.

77 Previous studies on synthetic TC generation have primarily employed two main ap-  
78 proaches: (1) statistical re-sampling, and (2) physical-based downscaling methods. Statisti-  
79 cal re-sampling models simulate TC genesis, tracks, and intensities (maximum wind speed  
80 or minimum central pressure) based purely on historical observational datasets, without  
81 considering environmental conditions. Examples of such models include those developed by  
82 (Vickery et al., 2000; James & Mason, 2005; Bloemendaal et al., 2020). These models typ-  
83 ically require a limited number of input variables, have low computational costs, and thus  
84 are easily applicable on a global scale. However, these models are not based on physical  
85 principles and cannot be accurately applied to a non-stationary climate due to changes in  
86 the background environment. On the other hand, physically-based downscaling methods  
87 relate TC characteristics to the large-scale background environmental conditions, making  
88 them environment-dependent. These methods can be statistical-deterministic (e.g., models  
89 developed by (K. Emanuel et al., 2006)) or purely statistical (e.g., models developed by  
90 (C.-Y. Lee et al., 2018; Jing & Lin, 2020)). Such environment-dependent approaches are  
91 capable of simulating the TC climatology in future climate scenarios and, therefore, are suit-  
92 able for climate change studies (K. Emanuel et al., 2008; C.-Y. Lee et al., 2020; Jing et al.,  
93 2021). Since the first synthetic TC downscaling approach in this family of models appeared  
94 in 2006, significant advances have been made in each of the three components (Huang et  
95 al., 2021; Huang, Wang, Jing, et al., 2022), and these approaches have been widely used  
96 for applications such as TC-induced surge risk assessment (Lin et al., 2012), regional loss  
97 assessment (Meiler et al., 2022; Huang, Wang, Liu, et al., 2022), and TC-induced wind load  
98 analysis (Kareem et al., 2019).

99 Both statistical and statistical-deterministic synthetic downscaling methods simulate  
100 the complete lifecycle of TCs using environmental parameters derived from the background  
101 environment. However, they cannot simultaneously simulate the two-way interactions be-  
102 tween the storm and its surrounding environment; therefore, the environment cannot re-  
103 spond correspondingly to the development of the storm. Furthermore, traditional ap-  
104 proaches do not comprehensively simulate the characteristics of TCs (i.e. genesis, track,  
105 intensity, and size) as a cohesive system. For example, although the TC intensity is deter-  
106 mined based on environmental predictors along the track, the storm track is predominantly  
107 driven by background winds, which is independent of the intensity component. Given the  
108 clear and strong correlation between these components, it is prudent to consider the poten-  
109 tial correlations between these components (Ruan & Wu, 2022). Moreover, the asymmetries  
110 in the TC surface wind field are not directly captured. Some approaches do not output the  
111 TC size and require a separate size component to determine the outer radius of the storm  
112 by random sampling from historical data (Jing & Lin, 2020). Other methods provide the  
113 radius of maximum wind (K. Emanuel et al., 2008); however, they require an additional  
114 parametric wind model to generate the full surface wind field, followed by post-processing  
115 to incorporate asymmetries related to storm translation speed and wind shear (Lin et al.,  
116 2012; Lin & Chavas, 2012).

117 The ideal synthetic downscaling method would simultaneously simulate all character-  
118 istics of the TC as an integrated system, including the interactions between storms and the  
119 surrounding environment, to generate detailed wind fields with greater accuracy but similar  
120 computational efficiency to that of traditional statistical and statistical-deterministic down-  
121 scaling methods. Recent progress in machine learning-based high-resolution global weather  
122 modeling (ML-GWM) (Pathak et al., 2022; Bi et al., 2022; Lam et al., 2022) has made  
123 this possible. ML-GWM systems are based on three-dimensional neural networks that are  
124 trained on high-quality reanalysis datasets, such as the ERA5 reanalysis (Hersbach et al.,  
125 2020), to predict weather around the globe. A significant advantage of ML-GWMs is their  
126 substantially lower computational costs compared to traditional numerical weather fore-

127 casting, while still operating at high spatial resolutions. Representing the cutting edge of  
 128 ML-GWM, Pangu-Weather (Bi et al., 2022) has outperformed the operational Integrated  
 129 Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts  
 130 in medium-range forecasting, with speeds more than 10,000 times faster. The high spa-  
 131 tial resolution of 0.25 degrees also enables Pangu-Weather to precisely track TCs based on  
 132 simulation results.

133 In this study, we leverage ML-GWM to create a novel ML-based approach for synthetic  
 134 TCs downscaling, which we call the synthetic TC-GENerative Model (or "TC-GEN"). This  
 135 involves generating a synthetic TC seed for each storm through a data-driven process, merg-  
 136 ing it with the background environment, and simulating both the storm and its surround-  
 137 ings simultaneously with Pangu-Weather. To achieve this, we first determine the annual  
 138 frequency, date, and location of synthetic TCs using an existing environment-based TC  
 139 genesis model. Next, we perform a Principal Component Analysis on all historical TCs at  
 140 genesis, identifying the principal components that effectively capture most of the variances  
 141 in TC genesis. Using these principal components, we generate synthetic TC seeds with  
 142 weights derived from historical data. We then integrate these TC seeds into the surround-  
 143 ing environment using Poisson blending, a technique widely used in image processing to  
 144 seamlessly merge two images, ensuring that the TC seeds are naturally embedded within  
 145 the larger environmental context while still maintaining important wind structures. Finally,  
 146 we run Pangu-Weather using their pre-trained model, which enables the joint simulation  
 147 of both the storm and its surrounding environment, and bias-correct simulated intensity to  
 148 real intensity. This integrated approach allows us to gather key characteristics of the TC,  
 149 such as the track and the maximum wind speed. It also provides spatial details such as  
 150 the full wind field, allowing for a direct derivation of the outermost extent of the storm.  
 151 It is worth noting that several key steps of this ML-based method are data driven, relying  
 152 heavily on historical data that lack substantial input from physics. Furthermore, while the  
 153 spatial resolution of 0.25 ° is a high resolution for global climate models, it is still too coarse  
 154 to accurately resolve the inner core and the structure of a storm. Therefore, we should use  
 155 the simulated three-dimensional structure including the horizontal surface wind field with  
 156 care, as it is likely to be unrealistic.

157 Despite these limitations, our ML-based method offers a unique set of collective ad-  
 158 vantages compared to previous TC downscaling approaches: 1) Holistic simulation: unlike  
 159 previous studies where the genesis, track, and intensity of the storm are simulated sep-  
 160 arately, this approach can simulate these three storm components together as a cohesive  
 161 system; 2) Integrated simulation: similar to numerical modeling, this approach can sim-  
 162 ulate both the storm and its environment simultaneously, thus allowing for the two-way  
 163 interactions between the storm and the environment; 3) Intrinsic asymmetry: while the TC  
 164 core and intensity may not be fully resolved, this approach has the capability to provide  
 165 crucial asymmetric characteristics in the TC surface wind field. This, in turn, allows for the  
 166 inference of the asymmetric outermost wind extent of a storm, which is essential for studying  
 167 tropical cyclone-induced hazards such as surges and heavy rainfall; 4) Efficiency: utilizing  
 168 pre-trained Pangu-Weather, this approach inherits the efficiency of statistical downscaling  
 169 methods that require little computational resources, enabling large samples of synthetic TCs  
 170 to be generated in days, a time frame comparable to the work of (Bloemendaal et al., 2020);  
 171 5) Scalability: this approach can be easily applied to other ocean basins and other high-  
 172 resolution climate datasets, including future climate projections such as those in CMIP6. A  
 173 successful extension of this approach is achieved when the climate dataset used for training  
 174 processes high resolution for detecting storm eyes, provides a reasonable representation of  
 175 the storm's outer size, and involves the development of a corresponding pre-trained machine  
 176 learning model as the core simulator.

177 To evaluate TC-GEN, we generate a large sample of synthetic TCs and compare those  
 178 simulated storms with historical observations. We also place TC-GEN in context with  
 179 an existing statistical-deterministic downscaling approach, represented by Emaneul et al.

180 (2008) (KE08, (K. Emanuel et al., 2008)) and a statistical downscaling method, represented  
 181 by the Princeton environment-dependent probabilistic model (PepC, (Jing & Lin, 2020)).  
 182 We choose these two existing approaches as they are both environment dependent and  
 183 have distinct model components. The metrics we use for comparison include the density of  
 184 the tracks over the ocean, the maximum lifetime intensity, the landfall frequency, and the  
 185 landfall intensity distributions. Given that KE08 and PepC lack the ability to simulate the  
 186 outer size of TC, we only compare the distribution of TC outer size simulated by TC-GEN  
 187 with that of the historical TCs identified using reanalysis data. In all of these metrics,  
 188 we demonstrate a strong alignment between simulated storms and observational data. We  
 189 further assess the adaptability of TC-GEN to different reanalysis datasets through two  
 190 analyses: one with the ERA5 reanalysis, on which Pangu-Weather is trained, but using  
 191 different temporal resolutions for model initialization, and the other using an alternative  
 192 reanalysis dataset from the National Centers for Environmental Prediction (NCEP). We  
 193 show that the effectiveness of TC-GEN depends on the consistency between the training  
 194 reanalysis dataset used to train the ML-GMW and the data used for model initialization.  
 195 Based on these analyses, we summarize the strength and limitations of TC-GEN and propose  
 196 potential improvements for future work.

## 197 **2 Pre-trained Models and Data**

### 198 **2.1 Neural network-based global weather model**

199 In this study, we use Pangu-Weather as the core ML-GWM to generate synthetic TCs.  
 200 Pangu-weather is an artificial intelligence-based model for medium-range global weather  
 201 forecasting, which has been shown to outperform the operational integrated forecasting  
 202 system of the European Center for Medium-Range Weather Forecasts (ECMWF) (Bi et al.,  
 203 2022).

204 Pangu-Weather is trained on atmospheric reanalysis data from the ERA5 reanalysis  
 205 data (Hersbach et al., 2020), using 69 atmospheric and surface variables as input and fore-  
 206 casting these variables for the subsequent time step. The 69 variables include 65 upper-air  
 207 variables (geopotential height, specific humidity, temperature, u and v component of wind at  
 208 13 pressure levels) plus four surface weather variables (2m temperature, u- and v- component  
 209 of 10m wind speed, and mean sea level pressure). Pangu-Weather offers four pre-trained  
 210 models that are capable of forecasting global weather data 1, 3, 6, and 24 hours ahead. We  
 211 use the model with a lead time of 6 hours since it is consistent with the temporal resolution  
 212 of TC IBTrACS data (see Section 2.2). Pangu-Weather has the ability to produce accurate  
 213 simulations that span multiple consecutive days, a time frame that is adequate to simulate  
 214 the entire life span of the most TCs. Pangu-Weather is open-source, and the pre-trained  
 215 models can be downloaded at <https://github.com/198808xc/Pangu-Weather>.

216 Pangu-Weather has several distinct advantages, making it highly suitable for this study.  
 217 First, Pangu-Weather is among the state-of-the-art ML-GWM systems with the highest  
 218 performance in weather forecasting. Second, Pangu-Weather has a high horizontal resolution  
 219 of  $0.25^\circ \times 0.25^\circ$ . Although the thermodynamic processes of the TCs cannot be fully resolved  
 220 at this spatial resolution, it is sufficient to detect the low pressure center, which enables  
 221 effective storm eye tracking. Third, Pangu-Weather is computationally efficient (more than  
 222 10,000 times faster than operational dynamical models), making it capable of simulating  
 223 large numbers of synthetic storms. In our experiments, we run Pangu-Weather with 8 RTX  
 224 A6000 GPUs, which adequately support all our computational tasks.

### 225 **2.2 Tropical Cyclone Data**

226 TC observations are extracted from IBTrACS data (version v04r00) (Knapp et al.,  
 227 2018, 2010), which includes the location (latitude and longitude) of each storm every 6  
 228 hours, along with its maximum sustained wind speeds measured at a height of 10 meters

229 above the sea surface. We use historical TC data in the North Atlantic Basin between 1979  
 230 and 2022 to generate TC seeds, correct biases in intensities simulated from Pangu-Weather,  
 231 and evaluate TC-GEN performance. We use extended TC data dating back to 1900 to  
 232 analyze landfall frequency, which provides a more precise assessment of sampling errors  
 233 with a larger number of TC tracks.

### 234 **2.3 Atmospheric Reanalysis Data**

235 Pangu-Weather requires global atmospheric and surface variables at the current time  
 236 step to forecast the next step. To be consistent with the input setup for the Pangu-Weather  
 237 model, we use ERA5 reanalysis with a resolution of  $0.25^\circ \times 0.25^\circ$ , the highest available  
 238 spatial resolution, and a temporal resolution of both 6-hour and monthly reanalysis data  
 239 to initiate Pangu-Weather. Using 6-hour reanalysis data for Pangu-Weather initialization  
 240 is intuitive, as it ensures alignment with the model setup and prevents domain gaps. How-  
 241 ever, generating synthetic TCs with 6-hourly reanalysis data presents two challenges. First,  
 242 introducing a synthetic TC to the 6-hourly reanalysis data may result in multiple storms  
 243 developing on the same map. This does not reflect reality, as the simultaneous occurrence  
 244 of multiple storms is relatively rare (Chowdhury et al., 2022). Second, generating synthetic  
 245 TCs by initiating Pangu-Weather with 6-hourly data requires downloading extensive his-  
 246 torical ERA5 reanalysis data, resulting in substantial data storage requirements. On the  
 247 contrary, starting Pangu-Weather with monthly data significantly reduces the data storage  
 248 burden and mitigates the problem of simultaneous occurrence. In this study, we use both  
 249 6-hourly and monthly reanalysis data for Pangu-Weather initialization, and under both sce-  
 250 narios we use the pre-trained model with a lead time of 6 hours to simulate storms. We  
 251 denote these two scenarios as 'TC-GEN hourly' and 'TC-GEN monthly' in the following  
 252 text and results.

253 In addition to ERA5 reanalysis, we test TC-GEN performance when initiating Pangu-  
 254 Weather with the NCEP GFS from Global Forecast Grids Historical Archive at  $0.25^\circ \times 0.25^\circ$   
 255 resolution (for Environmental Prediction/National Weather Service/NOAA/US Depart-  
 256 ment of Commerce, 2015). We opt for this reanalysis dataset because it provides all 69  
 257 variables required by Pangu-Weather, and it is available at a resolution of  $0.25^\circ \times 0.25^\circ$  with  
 258 global coverage.

### 259 **2.4 Existing downscaling approaches**

260 We contextualize TC-GEN by comparing it with two existing synthetic downscaling  
 261 approaches. The first is the statistical-deterministic model developed by Emanuel et al.  
 262 (2008) (K. Emanuel et al., 2008), which applies a random seeding method to initiate the  
 263 storm, a beta and advection model based on local winds to propagate the storm, and a  
 264 deterministic Coupled Hurricane Intensity Prediction System (CHIPS; (K. Emanuel et al.,  
 265 2004)) model to estimate the intensity of the storm based on the local thermodynamic state  
 266 of the atmosphere and ocean. The model has been extensively applied to assess TC hazards  
 267 (K. Emanuel, 2017; Marsooli et al., 2019), economic losses (Mendelsohn et al., 2012; Meiler  
 268 et al., 2022), and changes in TC climatology under future projected climate conditions  
 269 (K. Emanuel et al., 2008; Jing et al., 2021). Here, we compare our ML-based TC-GEN  
 270 results with a total of 4,100 synthetic TCs from 1980-2020 that intensify and reach TC  
 271 strength (lifetime maximum intensity exceeds 34 kt).

272 We also compare our TC-GEN results with the "PepC", a statistical synthetic down-  
 273 scaling approach of Jing and Lin (2020) (Jing & Lin, 2019). PepC has three components, a  
 274 genesis model, a track model, and an intensity model. The genesis component determines  
 275 annual frequency, as well as the time and location of weak vortices. The genesis model is de-  
 276 veloped using Poisson regression based on four large-scale environmental variables: absolute  
 277 vorticity, wind shear, relative humidity, and maximum potential intensity (K. A. Emanuel,  
 278 1988), which are contained in a genesis index (Tippett et al., 2011). After initialization,

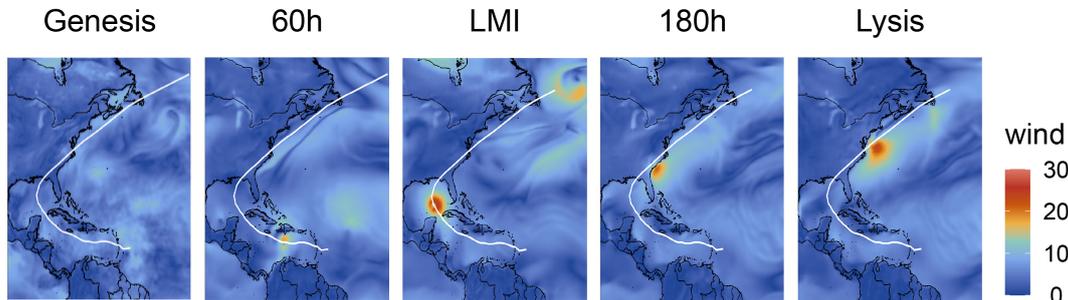
279 an analog-wind track model is used to propagate the storm based on analog factors (from  
 280 historical track patterns) and local in situ winds. The intensity of the storm is simulated as  
 281 a Markov process using an environment-dependent hurricane intensity model (Jing & Lin,  
 282 2019). In (Jing & Lin, 2020), the authors generated more than 55,000 TCs from 100 inde-  
 283 pendent realizations of the genesis of TCs during the period 1979-2014. To compare with  
 284 our ML-based TC-GEN results, we randomly select 10 independent 36-year realizations,  
 285 including a total of 3,690 TCs that intensify and reach TC strength.

### 286 **3 TC-GEN: data-driven synthetic TC downscaling approach**

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

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

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



**Figure 1. TC-GEN.** We propose an AI-empowered approach for synthetic TC downscaling where TC seed is created through a data-driven approach and then incorporated into a reanalysis environment map. Both the TC and its surrounding environment are then simulated in an interactive manner. The idea is illustrated in the figure, which presents five snapshots of total wind at different stages of the storm's life cycle.

#### 301 **3.1 Annual counts, date and locations of TC Seeds**

302 The first step of TC-GEN is to determine the annual frequency and the date and  
 303 location of the synthetic TCs. We use an environment-based hierarchical Poisson genesis  
 304 model to determine the number of synthetic storms that occur each year, as well as when  
 305 and where they originate over the ocean basin. This model is the genesis component of PepC  
 306 (Jing & Lin, 2020), which has been shown to generate TC formations that align closely with  
 307 observations, including interannual variations. This genesis model simulates the formation

of TCs for each month. When initializing Pangu-Weather with 6-hour reanalysis data, we randomly assign the precise date and time for each storm within the month of its genesis.

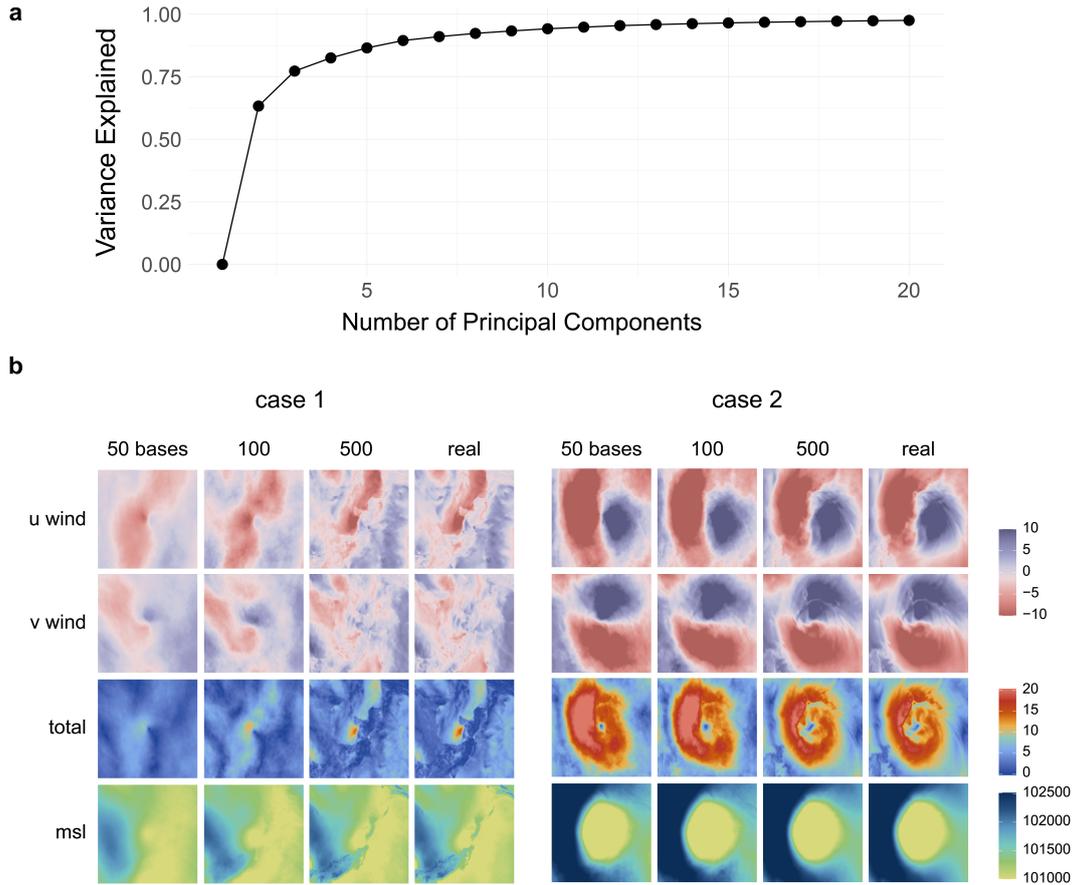
### 3.2 Structured TC seeds generation via data-driven approach

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 principal component analysis (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.

Principal Component Analysis (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 dataset. 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 (Nandi et al., 2016; Bretherton et al., 1992), and we refer the reader to the survey (Abdi & Williams, 2010) for more details.

We use 690 historical TCs in the period 1979 - 2022 to create 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 1600 kilometers), centered on the storm's eye. This circular region is large enough to encompass the outermost extent of most TCs at genesis. We collect TC seeds for each of the 69 atmospheric and surface variables from the ERA5 reanalysis for all  $M = 690$  storms, which form a matrix with dimensions of  $128 \times 128 \times 69 \times 690$ , representing historical TC seeds. Next, we perform PCA to reduce the dimensions of the data. The cumulative sum of the largest 20 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. To ensure quality, we use 500 principal components that represent the majority of the variations in actual environmental fields.

We create synthetic TC seeds by randomly sampling the low-dimensional space based on the top 500 principal components. 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. In Figure 3 we show three synthetic TCs at their genesis. We show four important variables (mean sea 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.

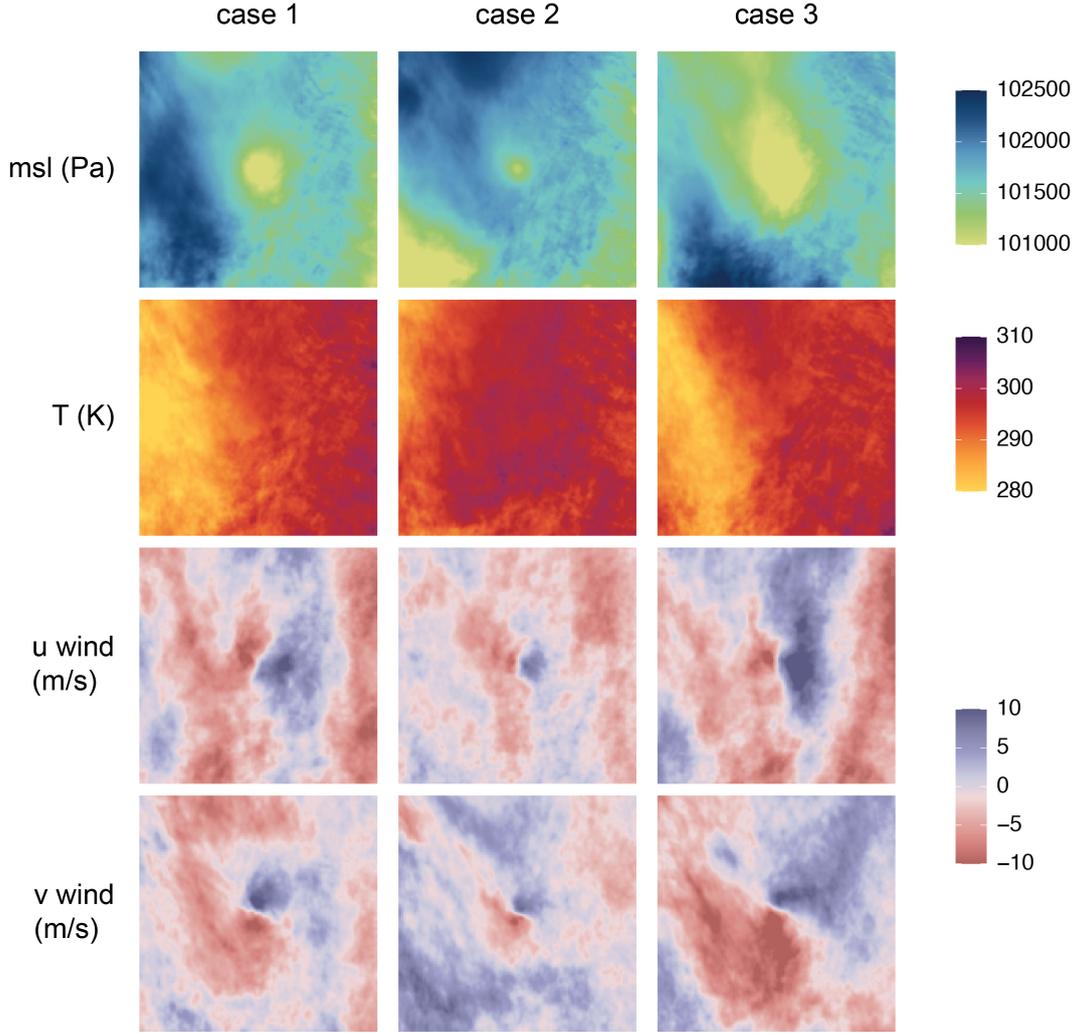


**Figure 2. Principal Component Analysis on TC genesis.** We run PCA on TC historical data to learn a low dimensional representation for TC seeds at genesis. Subplot (a) shows the cumulative sum of the eigenvalues sorted in descending order, where each point represents the proportion of total variance 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, using 50, 100, 500 principal components respectively. Both TC seeds are almost completely reconstructed using 500 principal components.

### 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 Sec 3.1). This step is critical for Pangu-Weather to simulate the environment along with the synthetic TC seeds. To achieve seamless integration of 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  $S \subset R^2$  (for example, color defined on pixel grids), and



**Figure 3. Synthetic TC Seeds.** This figure shows three synthetic TC seeds at genesis using the data-driven approach with 500 principal components. Each column shows one case, where mean sea level pressure, temperature, surface u- and v-winds are visualized.

$g$  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  $\Omega$ , such that

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

370 , where  $\nabla \cdot = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}]$ . This optimization aims to achieve a blended result where, within the  
 371 source region, it closely resembles the gradient of  $g$ , while at the boundary it should be  
 372 similar to  $f^*$ . For more details and numerical solutions, we refer the reader to (Pérez et al.,  
 373 2003).

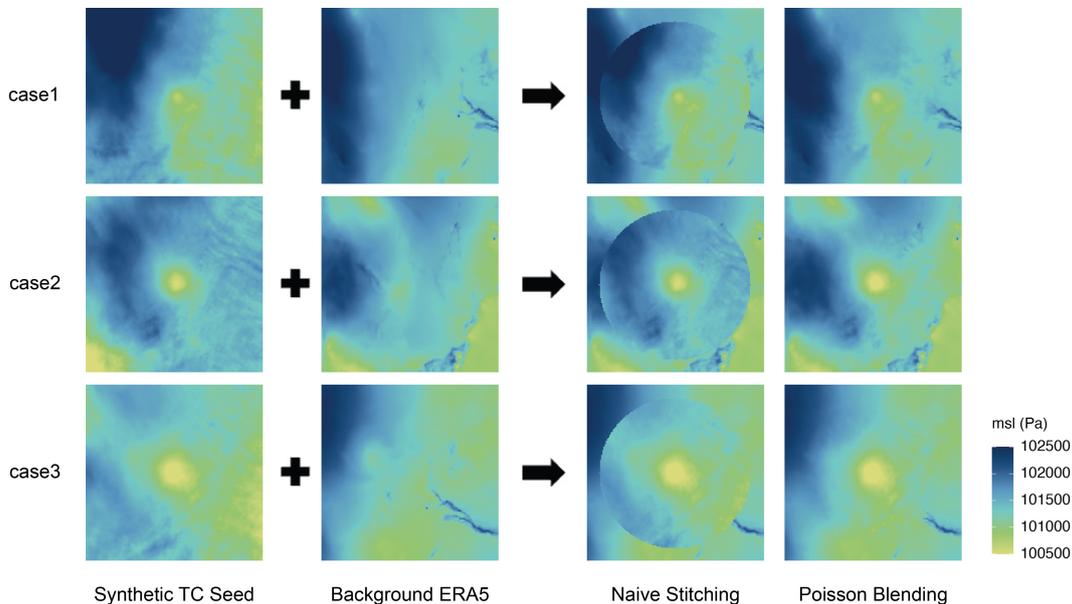
374 As an analogy to our problem,  $f^*$  and  $g$  represent the global environment map (that  
 375 is, target matrix) and synthetic TC seeds (that is, source matrix), respectively. In this  
 376 context, Poisson blending aims to combine these two matrices, ensuring that the gradients  
 377 of the blended results within the source region closely resemble the synthetic TC seeds, while  
 378 at the boundary, they should be similar to the background environment. In practice, we find  
 379 that naive Poisson blending still results in noticeable artifacts. Therefore, we adopt a linear

380 blending technique that incorporates both the source and target gradients, with a weight  
 381 determined by the distance to the boundary. The optimization is modified to accommodate  
 382 this approach:

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

383 where  $d$  is the distance of each pixel to the boundary and  $R$  represents the radius of the  
 384 blended region. This optimization guarantees a smooth integration of TC seeds into the  
 385 global environment map, with natural transitions at the boundaries and the structure of  
 386 TC well maintained.

387 We use this Poisson blending approach to seamlessly integrate the 69 atmospheric and  
 388 surface variables of synthetic TC seeds within a radius of 64 grid cells (approximately 1600  
 389 km) into the corresponding ERA5 reanalysis data. In Figure 4, we show the Poisson blend  
 390 process using mean sea level pressure as an example, which clearly reveals the eye of the  
 391 storm. Similarly, we show three cases as examples of weak, moderate, and strong storms,  
 392 illustrating the effectiveness of Poisson blending across a range of storm characteristics.



**Figure 4. 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 eye 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.

### 393 3.4 Simulation and tracking

394 After seamlessly integrating the generated TC seeds into the background environment  
 395 map, we proceed to run Pangu-Weather, simulating the entire life cycle of each storm over  
 396 a 15-day period, with 6-hour intervals between each step. To track the location of storm  
 397 center from Pangu-Weather outputs, we formulate a Gaussian kernel over mean sea level  
 398 pressure and vorticity at 10 meters, to identify the local maximum vorticity or minimum

399 pressure associated with the characteristic bell-shaped symmetric structure typical of a  
 400 storm. Then we determine the center of the storm by averaging the locations of these two  
 401 local extrema. Empirically, we set  $\sigma = 2$  for the Gaussian kernel and  $\sigma_1 = 2$ ,  $\sigma_2 = 8$  for  
 402 the Laplacian kernel, for the best performance. In most cases, the positions of maximum  
 403 vorticity closely align with those of minimum sea level pressure. However, when a TC’s  
 404 symmetric structure is not well maintained, this approach aids in stabilizing the outcomes  
 405 and provides robust tracking. It should be noted that Pangu-Weather includes an algorithm  
 406 for tracking TCs, which is based on relative vorticity, geopotential thickness, and 10-m wind  
 407 speed. Our tracking method delivers comparable results, while requiring fewer data inputs.  
 408 This advantage makes our approach generalizable to other weather forecasting systems that  
 409 may lack specific variables (for example, ForecastNet does not output vorticity; see (Pathak  
 410 et al., 2022)).

### 411 3.5 Bias correction of TC maximum wind speed

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. This leads to an underestimation of the maximum wind speeds of the TC. To address this problem, we develop a separate machine learning model to correct this bias based on the characteristics of the storm and the environment within the inner region of the storm. Due to the recurrent nature of this problem, where the wind speed at time  $t$  is highly correlated with the wind speed at previous time steps, we formulate the problem using a recurrent neural network based on long-short-term memory (LSTM). The structure of our network is shown in Figure 5 subplot b, where the bias correction stage is formulated as:

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

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

412 In Eq. 3, we use the term "raw" to indicate the intensity directly simulated from TC-  
 413 GEN, which is expected to have the same statistics as those in the ERA5 reanalysis data.  
 414 The term "true" is used to denote the real intensity, which has statistics similar to IBTrACs.  
 415 Thus,  $\mathbf{I}_{\text{raw}}$  and  $\mathbf{I}_{\text{true}}$  represent the maximum wind speed before and after bias correction.  
 416  $\mathbf{E}(\mathbf{x}(t))$  represents the environmental predictors extracted from the area surrounding the  
 417 TC center, which is located at  $\mathbf{x}(t)$  at time  $t$ .  $\mathbf{F}$  represents a machine learning model with  
 418 learnable weights  $\theta$ .

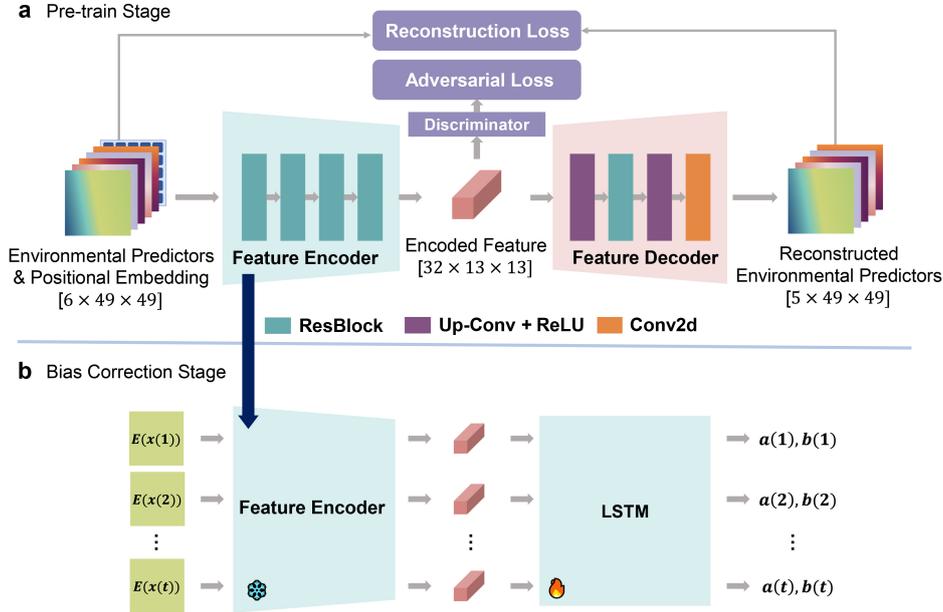
419 We use five environmental variables (mean sea level pressure, u- and v- component of  
 420 10m wind speed, relative humidity at 850 hpa and temperature at 850 hpa) that have been  
 421 identified as the most important predictors of the intensity of TC (Jing & Lin, 2019). At  
 422 each time step, the five variables within a circular region, covering a radius equivalent to  
 423 49 grid cells and centered at the TC, are combined with a positional embedding (Lam et  
 424 al., 2022) to create a raw input. We then convert the raw input to a feature vector using  
 425 a feature encoder comprised of 4 ResBlocks, and feed these feature vectors into an LSTM,  
 426 predicting  $a(t), b(t)$  at each time step.

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

433 To bridge the domain gap, we pre-train the feature encoder so that the extracted feature  
 434 is informative to reconstruct the raw ERA5 environment, yet indistinguishable in terms of  
 435 its source, i.e. whether it comes from ERA5 or Pangu-Weather. Such properties are achieved  
 436 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  
 438 by a feature decoder to reconstruct the original features using an L1 loss. Additionally,  
 439 we introduce an adversarial loss that performs a binary classification task, attempting to  
 440 discern the source of the feature (ERA5 or Pangu-Weather). The feature encoder is trained  
 441 to deceive the discriminator to the extent that it cannot identify the source of the feature.  
 442 As a result, the extracted features become domain-agnostic after this stage and the domain  
 443 gap is mitigated.

444 Overall, we start by training the auto-encoder architecture using a combination of  
 445 environment maps sampled from both the ERA5 and Pangu-Weather output (Figure 5  
 446 subplot a). Once the autoencoder converges, we discard the decoder, freeze the encoder  
 447 weights for feature extraction, and only update the LSTM weights (Figure 5 subplot b).  
 448 To train the LSTM, we divide the historical data of the TCs from 1979 to 2021, randomly  
 449 selecting 80% of the TCs for training and reserving the remaining 20% for testing. We use  
 450 the real maximum wind speed data from IBTrACS as ground truth. The LSTM is trained  
 451 with an AdamW optimizer (Loshchilov & Hutter, 2017), with a Huber loss and a learning  
 452 rate of 0.0001 over 10 iterations.

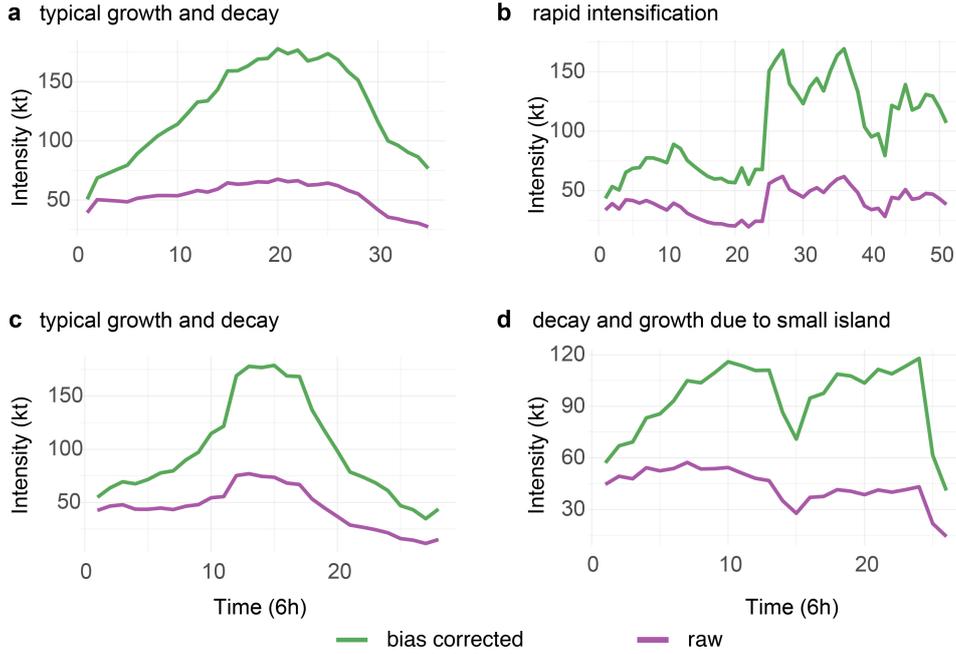


**Figure 5. Structure of the intensity bias correction model** The model consists of two stages, (a) a pre-trained stage to bridge domain gap between ERA5 and Pangu-Weather output so that all features become domain-agnostic; and (b) a bias correction stage that adjust the maximum wind speed using a Long Short-Term Memory (LSTM) model.

453 The performance of the LSTM-based bias correction model is illustrated in Figure 6,  
 454 which shows four cases that represent various characteristics of the storm. These include  
 455 typical growth and decay, rapid intensification, and storms that weaken after hitting an  
 456 island but subsequently strengthen after moving over the ocean.

### 457 3.6 Extracting the radius of outer size from each storm

458 The destructive potential of a TC is related to both its maximum sustained wind speed  
 459 and the radial extent of the near-surface wind, the latter typically measured by the outer



**Figure 6. Examples of bias correction in 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.

460 size of the TC (Powell & Reinhold, 2007; Irish et al., 2008). Therefore, it is important  
 461 to assess how well TC-GEN can replicate the realistic horizontal wind structures of TCs.  
 462 Furthermore, since TC seeds are artificially integrated into the surrounding environment,  
 463 successfully reproducing the historical distribution of the outer size of TCs can, to some  
 464 extent, prove that integration based on weighted Poisson blending is feasible.

465 Previous studies have shown that reanalysis datasets, including the ERA5 reanalysis  
 466 used in this study, are capable of effectively capturing the TC outer size, which is defined  
 467 by radii where convection is minimal and the atmosphere maintains a radiative–subsidence  
 468 balance (Schenkel et al., 2017). In earlier studies, multiple thresholds for azimuthal mean  
 469 tangential winds have been used, such as 2 m/s, 6 m/s, 8 m/s, 9 m/s, and 12 m/s (Bian et  
 470 al., 2021). Here, we use three size metrics based on the radii where the wind diminishes to  
 471 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  
 472 both the smallest and the largest metrics.

473 We identify the outer size of the TC following the procedures outlined in (Schenkel et  
 474 al., 2017). The procedure involves the following five main steps: 1) remove the background  
 475 flow from the surface wind field, which is empirically defined as the storm translation vector  
 476 rotated 20 deg cyclonically and diminished by a factor of 0.55, as described in the method  
 477 by (Lin et al., 2012); 2) establish a TC-centered polar coordinate system, where the cyclone  
 478 center serves as the origin, and the grid spacing in the radial direction is set at 0.5 times  
 479 the archived reanalysis grid spacing; 3) compute the tangential wind at the grid points of  
 480 the reanalysis data for each annular region of the polar coordinate system; 4) remove all  
 481 radial points if the data asymmetry parameter exceeds 0.5, following (Chavas & Vigh, 2014);  
 482 5) calculate the gridded radial profile of azimuthal-mean 10-m tangential wind using cubic  
 483 Hermite polynomial interpolation, based on which we then identify the radii of at which the

484 10-meter azimuthal-mean tangential wind speeds decrease to a certain threshold (12 m/s, 6  
485 m/s, and 2 m/s).

## 486 **4 TC-GEN Evaluation**

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

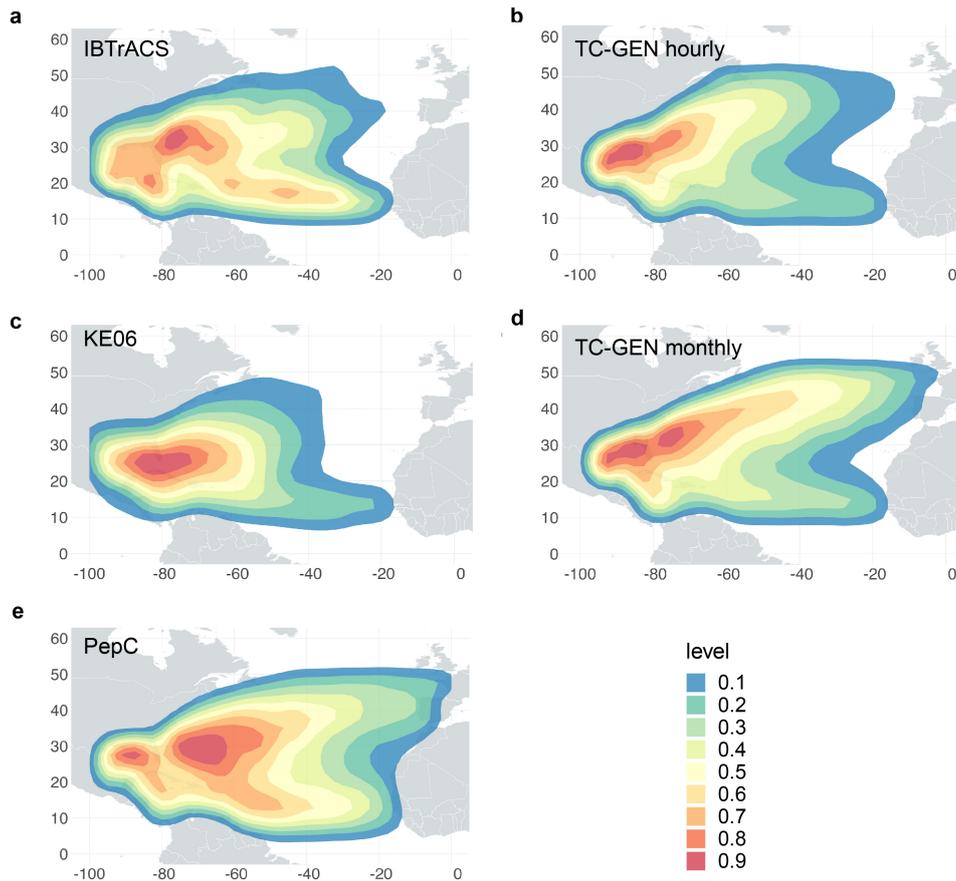
### 501 **4.1 Track Density**

502 Figure 7 compares the simulated tracks that are initiated with both hourly and monthly  
503 data, with observed tracks and simulated tracks using KE08 and PepC. The colors represent  
504 the spatial track density normalized to the maximum of the basin. We show that both sets of  
505 simulated tracks replicate the typical recurving pattern seen in the observed tracks relatively  
506 well, which is comparable to KE08 and PepC. Compared to observations, TC-GEN simulated  
507 tracks exhibit a negative bias in the main development region, mainly stemming from the  
508 negative bias in the genesis component of PepC within this region. We test this hypothesis  
509 with a sensitivity test by resampling the genesis according to the spatial distribution of  
510 historical genesis locations. After sampling, we find that the simulated tracks effectively  
511 capture the hotspots in the main development region, the southeast US coast, and the Gulf  
512 of Mexico, although there is a slight positive bias in the Gulf of Mexico with hourly data  
513 and in the West Caribbean Sea with monthly data (Figure S1).

514 We further evaluate the performance of TC-GEN by comparing the 6-hourly north-  
515 south and east-west displacements of simulated and observed tracks, which serves as a  
516 means to assess TC-GEN's performance in simulating individual tracks. The results are  
517 shown in Figure 8. All simulated data sets are largely consistent with the observations.  
518 For simulated tracks initiated with hourly data, there is a slight positive bias for positive  
519 meridional displacement, a slight negative bias for negative meridional displacement; and  
520 correspondingly, a negative bias for positive zonal displacements and a positive bias for  
521 negative zonal displacements, which could come from fewer recurvations in simulated storm  
522 tracks that originate in the main development region. Additionally, positive biases in merid-  
523 ional displacements may also arise from the deviations in eastward moving tracks at high  
524 latitude (near Europe), where they should have been terminated because of their low wind  
525 intensity. These patterns are also seen in simulated tracks that are initiated with monthly  
526 reanalysis data. In general, both datasets exhibit comparable performance to existing down-  
527 scaling methods, with simulated tracks initiated using hourly data showing slightly better  
528 performance than those initiated with monthly data.

### 529 **4.2 Landfall Frequency**

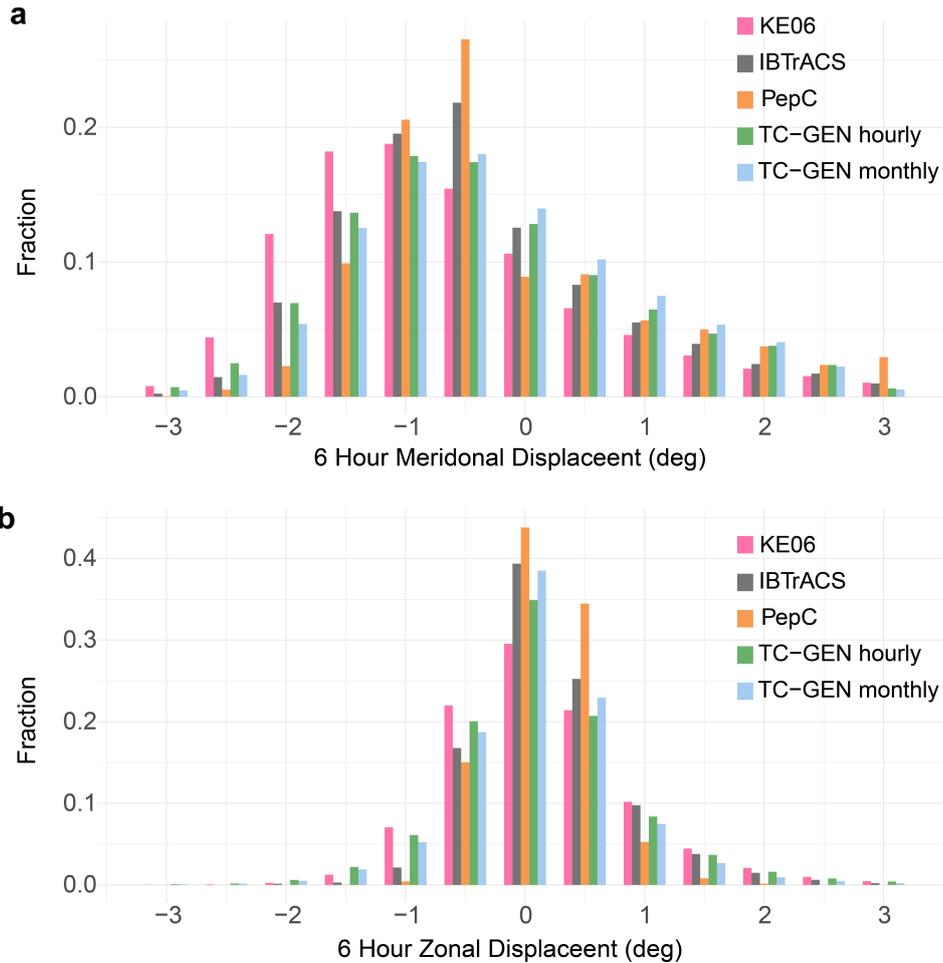
530 We examine the annual regional landfall frequency at coastal locations along the North  
531 Atlantic coastline. To help indicate locations, a total of 186 mileposts (MPs) are defined



**Figure 7. Track density** Track density is calculated as the accumulated number of TC passes into each  $0.75 \times 0.75$  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.

532 following Vickery et al. (2000) (Vickery et al., 2000), as shown in Figure 9, to cover the  
 533 coastline with 100-km spacing along the Mexican coastline and 50-km spacing along the US  
 534 coastline. Landfall is defined as simulated or observed storms that approach within 50 km  
 535 of each coastal milepost. The results presented in Figure 9b are based on simulated tracks  
 536 initiated using both hourly and monthly data. Furthermore, results from KE08- and PepC-  
 537 simulated tracks are included for comparison. As a reference, the historical annual landfall  
 538 rate between 1900 and 2022 is shown with shading that indicates the associated sampling  
 539 error at each milepost. The sampling error for the annual rate of each gate is determined  
 540 by calculating the total number and standard error of storms that cross each gate over the  
 541 entire record and then dividing both the mean and the error bars by the number of years.

542 Due to the different annual total frequencies in the simulated and observed track  
 543 datasets, which can significantly influence landfall frequency, we adjust the annual rate  
 544 to a uniform 13 storms per year in all datasets to form a fair comparison. Our results  
 545 indicate that the simulated tracks, from daily and monthly data, can reproduce the overall  
 546 pattern in the observations, showing a performance comparable to that of KE08 and PepC.  
 547 The annual landfall rates simulated by TC-GEN exhibit correlations with observations of  
 548 0.79 and 0.82 for the tracks initiated with hourly and monthly data, respectively, which are

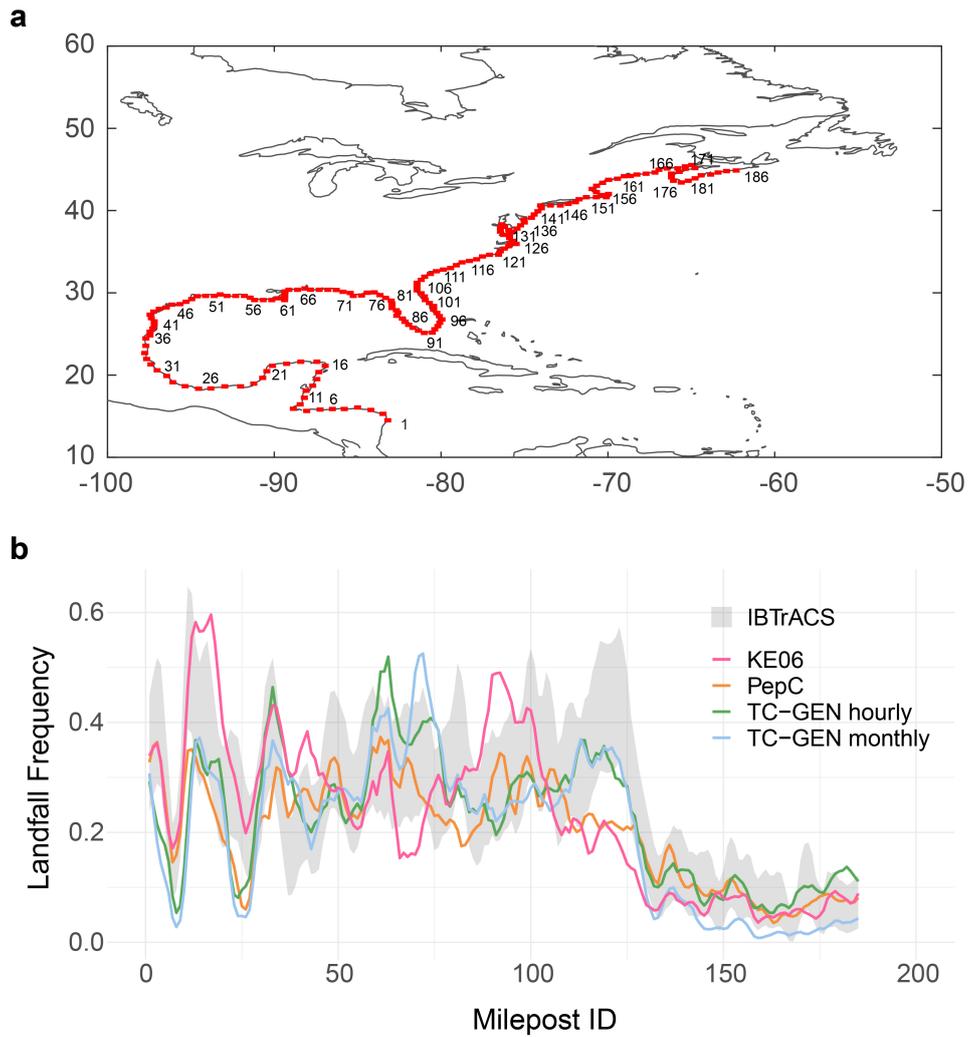


**Figure 8. 6 hour zonal and meridional displacement** Comparison of probability density functions of 6 hour (a) meridional and (b) zonal displacements between simulated tracks and observations. Simulated tracks initiated with hourly data and monthly data are both presented and compared with KE08 and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

549 comparable to 0.83 and 0.79 for PepC and KE08, respectively. Both TC-GEN simulated  
 550 data sets capture the landfall frequency at MP 100-125, which is overlooked by existing  
 551 methods, while they underestimate the landfall frequency below MP 25, possibly due to  
 552 fewer genesis in the main development region.

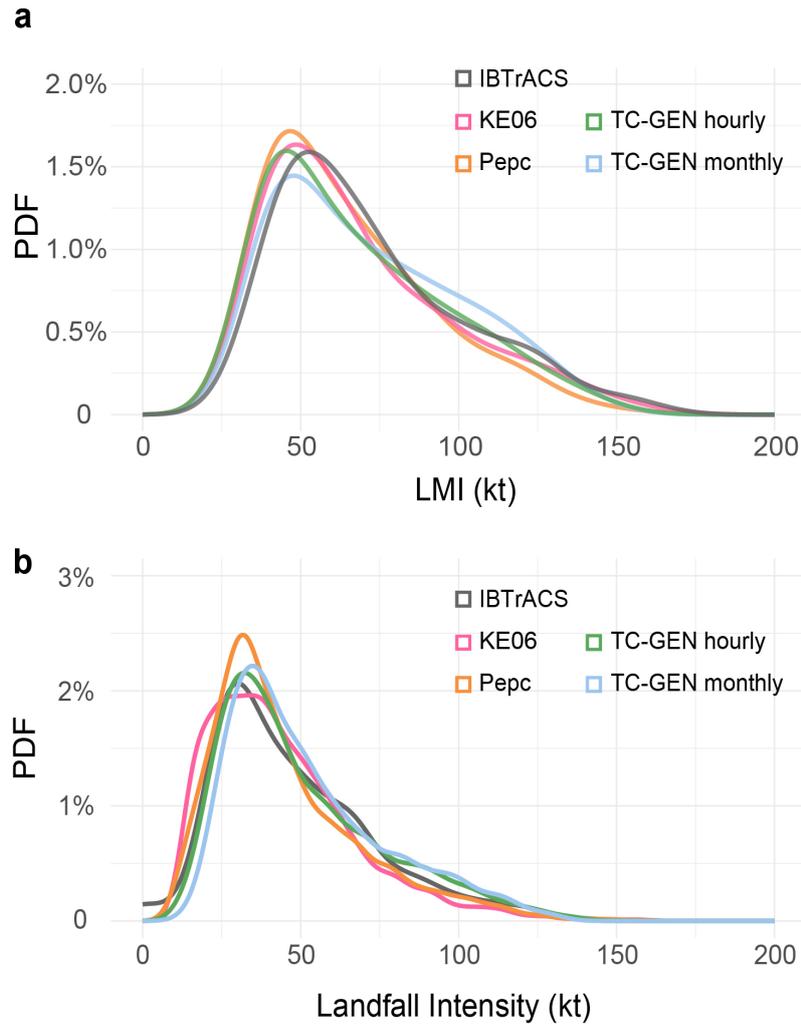
### 553 4.3 Lifetime Maximum Intensity and Landfall Intensity

554 We analyze the simulated TC intensity using two metrics, the lifetime maximum in-  
 555 tensity and landfall intensity. The LMI distribution serves as a representation of the TC  
 556 intensity climatology. A successful simulation should be able to reproduce the bimodal  
 557 distribution with a shoulder feature around 120 kt, which is associated with storms that  
 558 undergo rapid intensification (C.-Y. Lee et al., 2016). We show in Figure 10 that after bias  
 559 correction, both sets of TC-GEN are capable of reproducing a realistic distribution of the  
 560 observed LMI. TC-GEN tracks initiated with hourly data better simulate the tail of the LMI  
 561 distribution for LMI greater than 75 kt; however, there is a negative bias in LMI for storms  
 562 with LMI less than 75 kt, which are mostly moderate storms that do not undergo rapid



**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.

563 intensification. TC-GEN tracks initiated with monthly data overestimate the proportion of  
 564 storms with an LMI greater than 75 kt. The biases might originate from the intensity bias  
 565 correction model. In order to have a reasonable fraction of storms undergo rapid intensi-  
 566 fication, the bias correction model prioritizes strong storms, which could lead to a higher  
 567 proportion of storms becoming more intense than they should be. Similar patterns are also  
 568 seen in the distribution of the landfall intensity. Both data sets initiated with hourly and  
 569 monthly data exhibit a positive bias in landfall intensity greater than 75 knots. This bias is  
 570 also likely to be attributed to the intensity correction model, which tends to overestimate  
 571 the intensity of the storm when the storm has already weakened.



**Figure 10. Lifetime maximum intensity 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.

572

#### 4.4 Outer Size

573

574

575

576

577

578

579

580

581

582

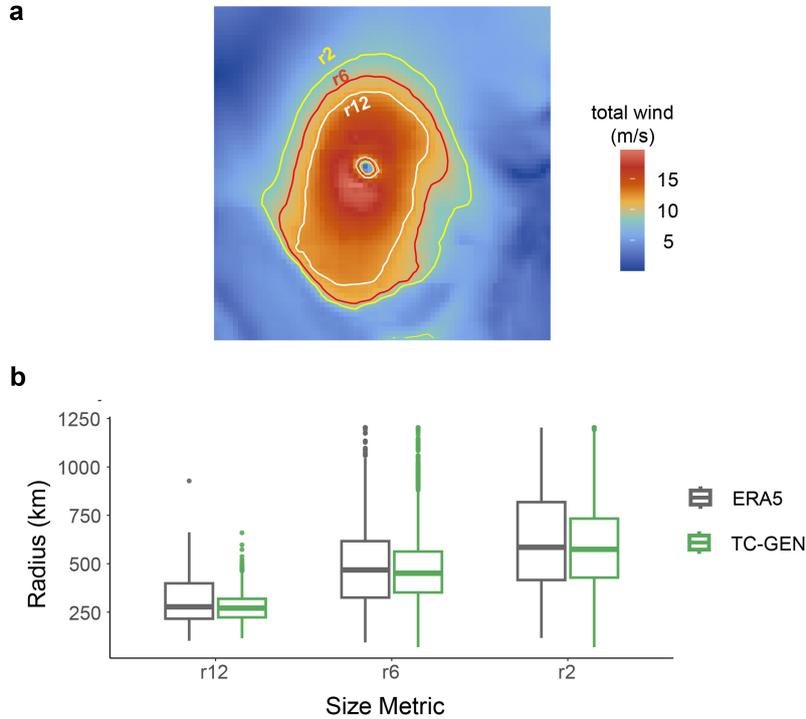
583

584

585

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, and ERA5 shows an improved representation of the outer size compared to previous versions of ERA (Bian et al., 2021). Here, we examine how the outer size of the TCs compares between TC-GEN simulated tracks and historical storms, where outer sizes of historical TCs are derived from the corresponding ERA5 reanalysis data. As the statistical downscaling approach does not provide this output, we do not have data from 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 outer size, with a discrepancy of approximately 10 km, which is even lower than the uncertainties in outer size arising from different reanalysis datasets (Bian et al., 2021). However, the

586 standard deviations for all three size metrics in TC-GEN-simulated TCs are smaller than  
 587 those observed in historical storms, particularly for r12, which represents the largest metric  
 588 that defines the outer size of storms. This smaller variation may arise from the bias in the  
 589 horizontal wind structures of TCs at genesis, which requires further investigation.



**Figure 11. 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-meter 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 ERA reanalysis. 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.

## 590 5 Discussion

### 591 5.1 Sensitivity to reanalysis data

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

598 Furthermore, we evaluate the suitability of applying TC-GEN to reanalysis data sets  
 599 other than ERA5 reanalysis. Using NCEP GFS (see 2.3, we find that the results are unsat-  
 600 isfactory (as illustrated in Figure S2), which is likely attributed to the well-known domain  
 601 gap problem in deep learning. In a deep learning model, typically the first several layers  
 602 of the neural network are responsible for extracting features from raw data. These layers  
 603 are trained to identify relevant features that help in downstream tasks, such as recogniz-

ing distinctive patterns, reducing noise, and achieving specific invariances. 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. Based on this discussion, an interesting future direction involves devising a generalized ML-based weather forecasting system that performs reasonably well on various reanalysis datasets. This could be accomplished, for example, by training the model on multiple reanalysis datasets 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. 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 200 km) to 89 grid cells (approximately 1200 km). Our sensitivity tests reveal that the results are relatively robust when the radius falls within the range of 49 to 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. In the development of this model, we also test 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. One limitation of the current bias correction model is that it tends to overestimate TC maximum intensity when the storm is decaying, which

partially explains the positive bias in the landfall intensity distribution. This bias 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). Consequently, the model tends to generate more storms that grow at a higher rate while decaying at a slower rate. Future work should focus on improving the bias correction model to effectively handle storms that undergo rapid intensification and those that do not, with the aim of achieving optimal performance in terms of both lifetime maximum intensity and landfall intensity.

## 5.5 Asymmetric TC wind field

Horizontal asymmetric TC wind fields are important for disaster management and regional risk assessment. In addition to the maximum wind speed, accurate estimation of the TC wind field is essential to identify at-risk populations and assess potential climate-related threats. One notable advantage of TC-GEN is its ability to directly output spatial asymmetric characteristics of the TC surface wind field. Although raw simulated winds are often underestimated as TC structures cannot be fully resolved at a resolution of  $0.25^\circ \times 0.25^\circ$ , there are still several ways to adjust the intensity of the wind and effectively use the simulated asymmetries. For example, with the adjusted maximum wind intensity and the extracted outer size of the TC, a parametric wind model can be applied, such as the model developed in Chavas et al. (2015) (Chavas et al., 2015), to generate the complete asymmetric wind profile of 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 a novel machine learning-based approach to the synthetic downscaling of tropical cyclones. This approach, which we refer to as "TC-GEN", leverages the recent advances in machine learning-based global weather models. The machine learning-based high-resolution global weather model (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 four key steps: 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, and the correction maximum wind speed biases using a long-short-term memory model. 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.

The recently introduced ML-GWM has significant potential for improvement with ongoing advances in machine learning and artificial intelligence, aiming for 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 tropical cyclone risk assessment, we believe that machine learning-based global weather models will play a crucial role in future climate studies.

## Open Research Section

Pangu-Weather trained models can be downloaded from the public GitHub repository at <https://github.com/198808xc/Pangu-Weather> (<https://doi.org/10.5281/zenodo.7678849>). Tropical cyclone observations are obtained from the International Best Track Archive for Climate Stewardship (IBTrACS) project at <https://www.ncei.noaa.gov/products/international-best-track-archive>. Historical ERA5 reanalysis data (both monthly and 6-hourly) are obtained from the ECMWF climate data store. Historical NCEP reanalysis data is downloaded at <https://rda.ucar.edu/datasets/ds084.1/>.

## Acknowledgments

K. Emanuel's research is part of the MIT Climate Grand Challenge on Weather and Climate Extremes. This research received support by the generosity of Eric and Wendy Schmidt by recommendation of Schmidt Futures as part of its Virtual Earth System Research Institute (VESRI).

## References

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4), 433–459.
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2022). Pangu-weather: A 3d high-resolution model for fast and accurate global weather forecast. *arXiv preprint arXiv:2211.02556*.
- 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.
- 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.
- 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.
- 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.
- Chavas, D. R., & Vigh, J. (2014). Qscat-r: The quikscat tropical cyclone radial structure dataset. *NCAR Tech. Note TN-5131STR*.
- 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. (2017). Assessing the present and future probability of hurricane harvey's rainfall. *Proceedings of the National Academy of Sciences*, 114(48), 12681–12684.
- Emanuel, K., DesAutels, C., Holloway, C., & Korty, R. (2004). Environmental control of tropical cyclone intensity. *Journal of the atmospheric sciences*, 61(7), 843–858.
- 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.
- Emanuel, K., Sundararajan, R., & Williams, J. (2008). Hurricanes and global warming:

- 752 Results from downscaling ipcc ar4 simulations. *Bulletin of the American Meteorological*  
753 *Society*, 89(3), 347–368.
- 754 Emanuel, K. A. (1988). The maximum intensity of hurricanes. *J. Atmos. Sci*, 45(7),  
755 1143–1155.
- 756 for Environmental Prediction/National Weather Service/NOAA/US Department of Com-  
757 merce, N. C. (2015). Ncep gfs 0.25 degree global forecast grids historical archive.  
758 *Research Data Archive at the National Center for Atmospheric Research, Computa-*  
759 *tional and Information Systems Laboratory.*
- 760 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., . . . oth-  
761 ers (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological*  
762 *Society*, 146(730), 1999–2049.
- 763 Huang, M., Wang, Q., Jing, R., Lou, W., Hong, Y., & Wang, L. (2022). Tropical cyclone  
764 full track simulation in the western north pacific based on random forests. *Journal of*  
765 *Wind Engineering and Industrial Aerodynamics*, 228, 105119.
- 766 Huang, M., Wang, Q., Li, Q., Jing, R., Lin, N., & Wang, L. (2021). Typhoon wind hazard  
767 estimation by full-track simulation with various wind intensity models. *Journal of*  
768 *Wind Engineering and Industrial Aerodynamics*, 218, 104792.
- 769 Huang, M., Wang, Q., Liu, M., Lin, N., Wang, Y., Jing, R., . . . Lou, W. (2022). Increasing  
770 typhoon impact and economic losses due to anthropogenic warming in southeast china.  
771 *Scientific reports*, 12(1), 14048.
- 772 Irish, J. L., Resio, D. T., & Ratchiff, J. J. (2008). The influence of storm size on hurricane  
773 surge. *Journal of Physical Oceanography*, 38(9), 2003–2013.
- 774 James, M., & Mason, L. (2005). Synthetic tropical cyclone database. *Journal of waterway,*  
775 *port, coastal, and ocean engineering*, 131(4), 181–192.
- 776 Jing, R., & Lin, N. (2019). Tropical cyclone intensity evolution modeled as a dependent  
777 hidden markov process. *Journal of Climate*, 32(22), 7837–7855.
- 778 Jing, R., & Lin, N. (2020). An environment-dependent probabilistic tropical cyclone model.  
779 *Journal of Advances in Modeling Earth Systems*, 12(3), e2019MS001975.
- 780 Jing, R., Lin, N., Emanuel, K., Vecchi, G., & Knutson, T. R. (2021). A comparison  
781 of tropical cyclone projections in a high-resolution global climate model and from  
782 downscaling by statistical and statistical-deterministic methods. *Journal of Climate*,  
783 34(23), 9349–9364.
- 784 Kareem, A., Hu, L., Guo, Y., & Kwon, D.-K. (2019). Generalized wind loading chain: Time-  
785 frequency modeling framework for nonstationary wind effects on structures. *Journal*  
786 *of Structural Engineering*, 145(10), 04019092.
- 787 Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, M. C., & Schreck, C. J. (2018). Interna-  
788 tional best track archive for climate stewardship (ibtracs) project, version 4.v04r00.  
789 doi: doi:10.25921/82ty-9e16
- 790 Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010).  
791 The international best track archive for climate stewardship (ibtracs) unifying tropical  
792 cyclone data. *Bulletin of the American Meteorological Society*, 91(3), 363–376.
- 793 Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., . . .  
794 others (2022). Graphcast: Learning skillful medium-range global weather forecasting.  
795 *arXiv preprint arXiv:2212.12794*.
- 796 Lee, C.-Y., Camargo, S. J., Sobel, A. H., & Tippett, M. K. (2020). Statistical–dynamical  
797 downscaling projections of tropical cyclone activity in a warming climate: Two di-  
798 verging genesis scenarios. *Journal of Climate*, 33(11), 4815–4834.
- 799 Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2016). Rapid intensification  
800 and the bimodal distribution of tropical cyclone intensity. *Nature communications*,  
801 7(1), 10625.
- 802 Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2018). An environmentally  
803 forced tropical cyclone hazard model. *Journal of Advances in Modeling Earth Systems*,  
804 10(1), 223–241.
- 805 Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., . . . others (2023).  
806 Climate change 2023: synthesis report. contribution of working groups i, ii and iii to

- 807 the sixth assessment report of the intergovernmental panel on climate change.
- 808 Lin, N., & Chavas, D. (2012). On hurricane parametric wind and applications in storm
- 809 surge modeling. *Journal of Geophysical Research: Atmospheres*, *117*(D9).
- 810 Lin, N., Emanuel, K., Oppenheimer, M., & Vanmarcke, E. (2012). Physically based assess-
- 811 ment of hurricane surge threat under climate change. *Nature Climate Change*, *2*(6),
- 812 462–467.
- 813 Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint*
- 814 *arXiv:1711.05101*.
- 815 Marsooli, R., Lin, N., Emanuel, K., & Feng, K. (2019). Climate change exacerbates hurricane
- 816 flood hazards along us atlantic and gulf coasts in spatially varying patterns. *Nature*
- 817 *communications*, *10*(1), 3785.
- 818 Meiler, S., Vogt, T., Bloemendaal, N., Ciullo, A., Lee, C.-Y., Camargo, S. J., ... Bresch,
- 819 D. N. (2022). Intercomparison of regional loss estimates from global synthetic tropical
- 820 cyclone models. *Nature Communications*, *13*(1), 6156.
- 821 Mendelsohn, R., Emanuel, K., Chonabayashi, S., & Bakkensen, L. (2012). The impact
- 822 of climate change on global tropical cyclone damage. *Nature climate change*, *2*(3),
- 823 205–209.
- 824 Nam, H., Lee, H., Park, J., Yoon, W., & Yoo, D. (2021). Reducing domain gap by reducing
- 825 style bias. In *Proceedings of the ieee/cvf conference on computer vision and pattern*
- 826 *recognition* (pp. 8690–8699).
- 827 Nandi, A., Mandal, A., Wilson, M., & Smith, D. (2016). Flood hazard mapping in jamaica
- 828 using principal component analysis and logistic regression. *Environmental Earth Sci-*
- 829 *ences*, *75*, 1–16.
- 830 Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., ...
- 831 Anandkumar, A. (2022). Fourcastnet: A global data-driven high-resolution weather
- 832 model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*.
- 833 Pérez, P., Gangnet, M., & Blake, A. (2003). Poisson image editing. *ACM SIGGRAPH 2003*
- 834 *Papers*.
- 835 Pörtner, H.-O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., ... others (2022).
- 836 Climate change 2022: Impacts, adaptation and vulnerability. *IPCC Sixth Assessment*
- 837 *Report*.
- 838 Powell, M. D., & Reinhold, T. A. (2007). Tropical cyclone destructive potential by integrated
- 839 kinetic energy. *Bulletin of the American Meteorological Society*, *88*(4), 513–526.
- 840 Ruan, Z., & Wu, Q. (2022). Relationship between size and intensity in north atlantic
- 841 tropical cyclones with steady radii of maximum wind. *Geophysical Research Letters*,
- 842 *49*(3), e2021GL095632.
- 843 Schenkel, B. A., Lin, N., Chavas, D., Oppenheimer, M., & Brammer, A. (2017). Evaluating
- 844 outer tropical cyclone size in reanalysis datasets using quikscat data. *Journal of*
- 845 *Climate*, *30*(21), 8745–8762.
- 846 Tippett, M. K., Camargo, S. J., & Sobel, A. H. (2011). A poisson regression index for
- 847 tropical cyclone genesis and the role of large-scale vorticity in genesis. *Journal of*
- 848 *Climate*, *24*(9), 2335–2357.
- 849 Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., ... Birchfield,
- 850 S. (2018). Training deep networks with synthetic data: Bridging the reality gap by
- 851 domain randomization. In *Proceedings of the ieee conference on computer vision and*
- 852 *pattern recognition workshops* (pp. 969–977).
- 853 Vickery, P., Skerlj, P., & Twisdale, L. (2000). Simulation of hurricane risk in the us using
- 854 empirical track model. *Journal of structural engineering*, *126*(10), 1222–1237.
- 855 Wei, L., Zhang, S., Gao, W., & Tian, Q. (2018). Person transfer gan to bridge domain gap
- 856 for person re-identification. In *Proceedings of the ieee conference on computer vision*
- 857 *and pattern recognition* (pp. 79–88).

# Supporting Information for ”TC-GEN: Data-driven Tropical Cyclone Downscaling using Machine Learning-Based High-resolution Weather Model”

Renzhi Jing <sup>1,2</sup>, Jianxiong Gao <sup>3</sup>, Yunuo Cai <sup>4</sup>, Dazhi Xi <sup>5</sup>, Yinda Zhang <sup>6</sup>,

Yanwei Fu <sup>4</sup>, Kerry Emanuel <sup>7</sup>, Noah S. Diffenbaugh <sup>1,8,9</sup>, Eran Bendavid <sup>2,10</sup>

<sup>1</sup>Woods Institute for the Environment, Stanford University, Stanford, CA, USA

<sup>2</sup>School of Medicine, Stanford University, Stanford, CA, USA

<sup>3</sup>Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai, China

<sup>4</sup>School of Data Science and MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China

<sup>5</sup>Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA

<sup>6</sup>Google LLC, 1600 Amphitheatre Pkwy, Mountain View, CA, USA

<sup>7</sup>Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>8</sup>Department of Earth System Science, Stanford University, Stanford, CA, USA

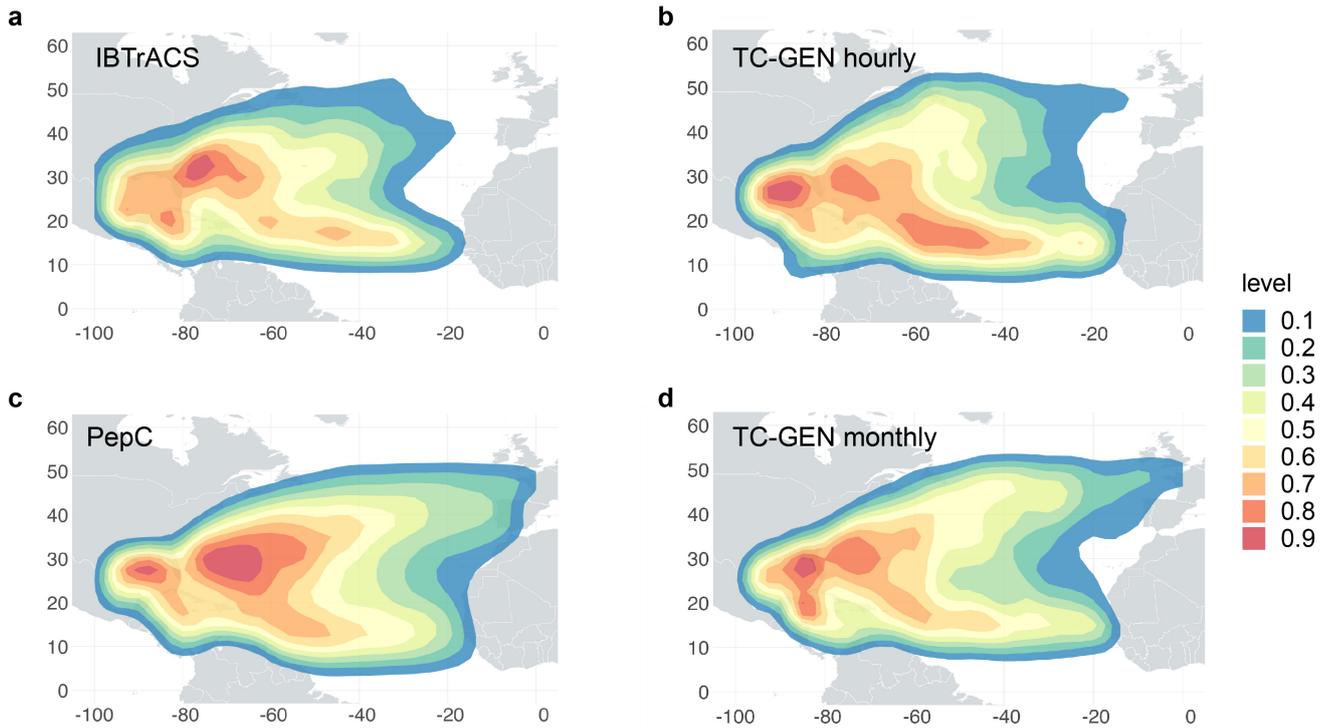
<sup>9</sup>Doerr School of Sustainability, Stanford University, Stanford, CA, USA

<sup>10</sup>Department of Health Policy, Stanford University, Stanford, CA, USA

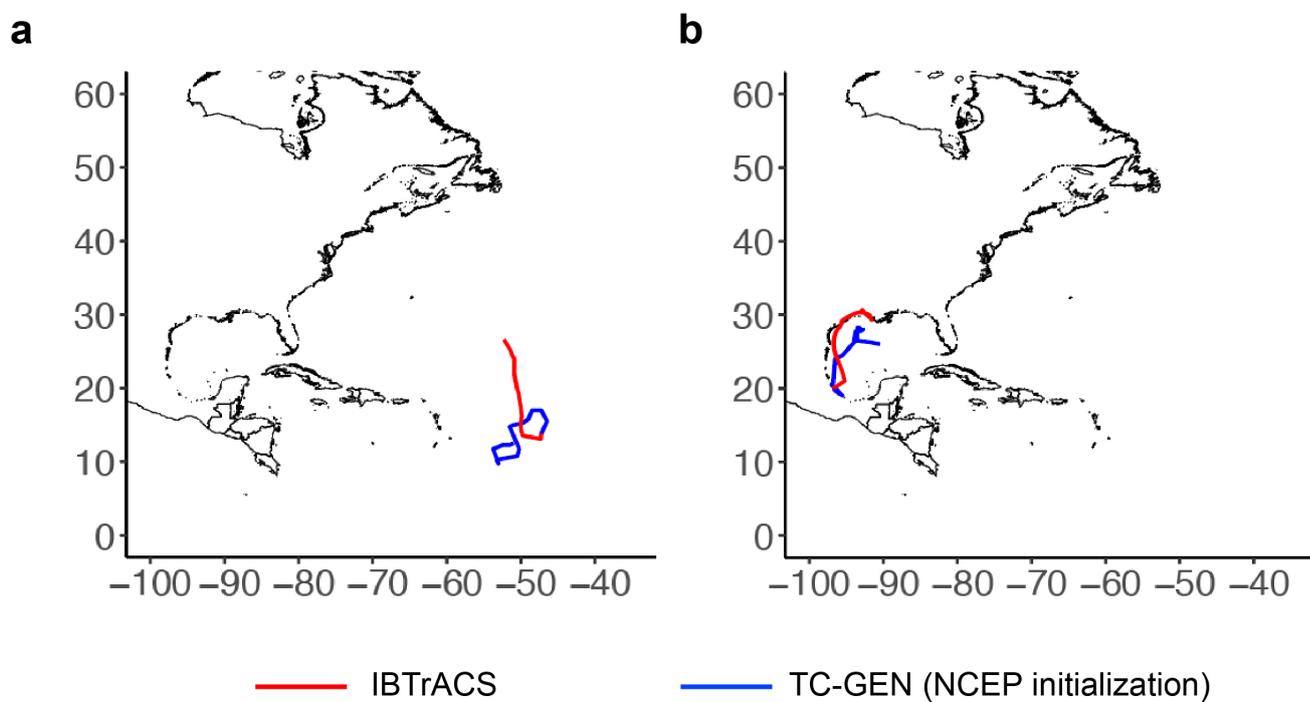
## Contents of this file

1. Figures S1 to S2

---



**Figure S1. Sensitivity test of track density** Similar to Figure 7, track density is computed as the accumulated number of tropical cyclone passes into each  $0.75 \times 0.75$  grid box, normalized by the maximum grid value of the basin, and smoothed using a Gaussian low-pass filter. The four panels show track density from (a) observations, (b) TC-GEN initiated with hourly data, (c) storms simulated by PepC, and (d) TC-GEN initiated with monthly data. In comparison with Figure 7, we resample the genesis locations based on the spatial distribution of historical genesis to ensure a fair comparison between TC-GEN simulated storms and observations. We demonstrate that, following the resampling, the simulated tracks effectively capture the hotspots in the main development region, the southeast US coast, and the Gulf of Mexico.



**Figure S2. Robustness with alternative reanalysis data** We assess the suitability of TC-GEN by initiating it with an alternative reanalysis dataset, the NCEP GFS. Through the simulation of two historical tropical cyclones (a) 2021240N13313 and (b) 2021256N21265, representing storms originating in the Gulf of Mexico and the main development region, respectively, we observe unsatisfactory performance. This is likely due to the well-known domain gap problem in deep learning.