

1 **An environmentally forced tropical cyclone hazard model**

2 Chia-Ying Lee*

3 *International Research Institute for Climate and Society, Columbia University, Palisades, NY*

4 Michael K. Tippett

5 *Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY*

6 *Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia*

7 Adam H. Sobel

8 *Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY*

9 *Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY*

10 Suzana J. Camargo

11 *Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY*

12 *Corresponding author address: Chia-Ying Lee, International Research Institute for Climate and
13 Society, Columbia University, 61 Route, 9W, Palisades, NY 10964.

14 E-mail: clee@iri.columbia.edu

ABSTRACT

15 A physics-based statistical stochastic system is developed for estimating the
16 long-term hazard of rare, high impact landfall events globally from ensem-
17 bles of synthetic tropical cyclones. There are three components representing
18 the complete storm lifetime: an index-based genesis model, a beta-advection
19 track model and an autoregressive intensity model. All three components de-
20 pend upon the local environmental conditions, including potential intensity,
21 relative sea surface temperature, 850 and 250 hPa steering flow, deep-layer
22 mean vertical shear, 850 hPa vorticity, and midlevel relative humidity. The
23 hazard model, using 400 realizations of a 32-year period (approximately 3000
24 storms per realization), captures many aspects of tropical cyclone statistics,
25 such as genesis and track density distribution. Of particular note, it simu-
26 lates the observed number of rapidly intensifying storms, a challenging issue
27 in tropical cyclone modeling and prediction. Using the return period curve
28 of landfall intensity as a measure of local tropical cyclone hazard, the model
29 reasonably simulates the hazard in the western north Pacific (coastal regions
30 of the Philippines, China, Taiwan, and Japan) and the Caribbean islands. In
31 other regions, the observed return period curve can be captured considering a
32 local landfall frequency adjustment.

33 **1. Introduction**

34 From 1963 to 2012, tropical cyclones (TCs) were responsible for more than 50% of all
35 meteorologically-induced financial losses (Geiger et al. 2016). TC hazard assessment is impor-
36 tant to government, industry, finances, NGOs, and even individual households in the context of
37 individual events, seasonal predictions, and climate adaptation. Accurate risk assessment depends
38 on the hazard — the probability of a TC of a given magnitude in a given location — in addition to
39 vulnerability factors, such as the growth of wealth and population (Estrada et al. 2015). We focus
40 on hazard in this study. Because of the limited historical record, a common approach for estimating
41 TC hazard is to compute statistics from simulated as well as observed storms (e.g., Emanuel et al.
42 2008). In this approach, each the complete lifetime of each simulated storm, including its genesis,
43 track, intensity and landfall, are simulated. Alternatively, one can statistically model the landfall
44 rate alone (Tolwinski-Ward 2015). Most industry catastrophe models (models which represent TC
45 hazard as well as vulnerability and financial losses to insured assets) use statistical methods to gen-
46 erate synthetic storms that are similar to those in historical data (e.g., AIR WORLDWIDE 2015).
47 Some of them include the dependence of storm activity on a few environmental parameters, such
48 as basin sea surface temperature (SST) or measures of the El Niño–Southern Oscillation (ENSO)
49 (e.g., Hall and Jewson 2007; Yonekura and Hall 2011, 2014). These models, while they generally
50 perform well in the current climate, are strongly constrained to the historical records and are not
51 designed to consider the effects of climate change.

52 To understand the impact of climate change on TC hazard, global climate models or dynamical
53 downscaling methods are the most straightforward approaches. Such models calculate individ-
54 ual TC evolution based on the laws of physics, and can provide information globally (whereas
55 many statistical models are developed for individual basins). However, at the high spatial resolu-

56 tions necessary for TC simulation, it is computationally expensive to generate a sufficient number
57 of synthetic storms for hazard assessment, where one is particularly interested in very rare and
58 extreme events. Thus, Emanuel et al. (2006) proposed a novel statistical–dynamical downscal-
59 ing method. In this method, each TC’s evolution is calculated using a combination of statistical
60 and simplified dynamical models that are forced by environmental conditions taken from global
61 models. The model of Emanuel et al. (2008) randomly seeds storms globally, moves them us-
62 ing a beta-advection model (Marks 1992), and calculates intensity evolution using a simple cou-
63 pled ocean–atmosphere tropical cyclone model (CHIPS, Emanuel et al. 2004). Emanuel’s model
64 has been broadly used for understanding the impact of a changing climate on TC climatology
65 (Emanuel 2013, 2015), storm surge hazard (Lin et al. 2012), and TC-induced economic losses
66 (Geiger et al. 2016).

67 In previous work we focused on developing a model for the intensity evolution, which is a
68 challenging issue even for hurricane forecasting. Lee et al. (2015, 2016a) describe a global au-
69 toregressive (AR) TC intensity model. The AR model contains a deterministic component derived
70 empirically, which advances the TC intensity in time and accounts for the surrounding large-scale
71 environment. The stochastic forcing of the AR model represents the component of TC intensi-
72 fication that is not linearly related to the storm’s ambient conditions. Simulating the intensity
73 evolution along the observed tracks, the AR model captures the observed TC intensity climatol-
74 ogy well, except for the bimodal distribution in LMI, which is associated with rapid intensification,
75 and is important for the simulating the frequency of the most intense storms (Lee et al. 2016b). In
76 this study, we will show that the AR intensity model is capable of simulating the observed LMI
77 distribution when the simulated storm lifetime is determined consistently with the intensity model,
78 rather than by the lifetime of the prescribed tracks. Other intensification models include that of
79 Lin et al. (2017), who used a multiple linear regression model, found that the dependence of TC

80 intensification to environment is nonhomogeneous and suggested a mixture modeling approach
81 as a solution. Recently, Emanuel (2017) reduced the complicity of his intensity model to a set
82 of two prognostic equations for storm intensity and inner-core moisture and further increased the
83 efficiency of his hazard model.

84 In the present study, we develop and assess a complete statistical-dynamical downscaling TC
85 hazard model. We develop new genesis and track components and couple them to the existing
86 AR intensity model described in detail in Lee et al. (2015, 2016a). Both the genesis and track
87 components depend on the local environment. Thus, the whole system is environmentally forced
88 with no explicit spatially dependent component. The model is developed for current climate with
89 all the environmental parameters downscaled from ERA-Interim. The data and methods used
90 for the model development and evaluation are described in Section 2. We introduce the individual
91 model components (genesis, track, and intensity), respectively, in Section 3. The TC hazard model
92 performance is first evaluated by its ability to capture the observed TC climatology, including fre-
93 quency, intensity, landfall, and interannual variability (Section 4). Next, we compare the observed
94 and simulated hazard in various places across global (Section 5). Throughout this study, we de-
95 fine ‘hazard’ as the probability (or equivalently the return period) of the storm intensity at landfall
96 exceeding a given threshold at a particular location. The summary and discussion are given in
97 Section 6.

98 **2. Data and Methods**

99 *a. Observational and reanalysis datasets*

100 The best-track dataset, HURDAT2, produced by the National Hurricane Center (NHC) is used
101 for the North Atlantic (ATL) and Eastern North Pacific (ENP) (Landsea and Franklin 2013; NHC

102 2013). For TCs in the Western North Pacific (WNP), Indian Ocean (IO), and Southern Hemisphere
103 Ocean (SH), we use the best-track data from Joint Typhoon Warning Center (JTWC, Chu et al.
104 2002; JTWC 2014). Both datasets include 1-min maximum sustained wind, minimum sea level
105 pressure, and storm location every 6 hours. Large-scale environmental variables are calculated
106 from the European Centre for Medium-Range Weather Forecasts interim reanalysis (ERA-Interim,
107 Dee et al. 2011; ECMWF 2013). We use monthly data for all three model components. In the track
108 model, additional daily 250 and 850 hPa steering flow winds are used as well. In this study, data
109 from 1981 to 2012 are used for evaluation. Data from 1981 to 1999 are used as training data for
110 the intensity model¹.

111 Throughout this study, the Saffir–Simpson scale is used to categorize storm strength in all basins.
112 The ranges used are 64–82 kt for category 1 (Cat 1) storms, and 83–95, 96–112, 113–136, >137 kt
113 for categories 2-5 (Cat 2-5) storms, respectively. The threshold for tropical storm (TS) is 34 kt.
114 Storm lifetime maximum intensity (LMI) is defined as the maximum sustained wind speed during
115 the storm’s life cycle.

116 *b. Identifying landfall locations*

117 For risk assessment, it is important to calculate the landfall probability at a given location, and
118 thus to identify landfall. We first linearly interpolate track data (for both observations and simu-
119 lations) to a 15-minute resolution. Surface type (land or ocean) is assigned to each interpolated
120 point using 0.5-degree resolution topography data from NASA (https://neo.sci.gsfc.nasa.gov/view.php?datasetId=SRTM_RAMP2_TOPO). Then, landfall is defined when a storm center
121 moves from a ocean point to a land point. To avoid counting landfalls multiple times in the situa-
122

¹The training of the intensity model was done in previous study (Lee et al. 2016a), and therefore was using data from different period.

123 tion when a storm moves over archipelago regions, such as the Philippines, landfalls need to be at
124 least 100 km and 6 hours apart to be considered as independent landfalls.

125 *c. Experimental design*

126 Simulations from the TC hazard model will be called GTI here, in which ‘G’, ‘T’, and ‘I’
127 stand for Genesis, Track, and Intensity models, respectively. In order to isolate the influence of
128 the individual components on the estimated TC statistics and hazard, we design two additional
129 experiments: $\hat{G}\hat{T}\hat{I}$ uses only the intensity model along with the genesis and tracks, represented as
130 ($\hat{\cdot}$), from the best-track dataset; $\hat{G}TI$ uses both track and intensity models, but observed genesis
131 locations. As we will discuss in the next section, each of three components in the TC hazard model
132 contains a stochastic parameter. Thus, the hazard model is a stochastic system. We construct 400
133 realizations of a 32 year period (1981 to 1999) in every experiment. In $\hat{G}\hat{T}\hat{I}$, the 400 realizations
134 differ in only in the component due to the intensity model. In $\hat{G}TI$, there are 10 sets of tracks
135 (with the same observed genesis locations) and each set has 40 intensity realizations. Realizations
136 with the same underlying tracks but different intensities can still differ in their lifetimes (due to
137 the different realizations of the intensity model solution), and thus in how much of each track is
138 actually covered by a storm. A similar design is used for GTI but the genesis locations in each set
139 are calculated from the genesis model separately.

140 *d. Evaluation measures*

141 To evaluate a stochastic model performance, we use two statistical measures:

142 The **Z-score** of a variable is defined as the observed minus simulated ensemble mean divided by
143 the observed variance. In an unbiased model, the Z-score magnitude should be smaller than one in
144 most areas, because the model error is small compared to the natural variability. The distribution of

145 Z-score also tells whether the bias is systematic (i.e., has a pattern) or nonsystematic (the positive
146 and negative values are randomly distributed).

147 The **Rank histogram** of a variable is defined as the distribution of the rank (in percentage) of
148 the observations with respect to the simulations. If the ensemble members and the observations
149 are drawn from the same probability distribution, the rank of observations with respect to the
150 simulations will be uniformly distributed. When the simulation is biased, under- or over-dispersed,
151 the shape of rank histogram will be tilted, bimodal with peaks at two ends, or mono-modal.

152 **3. Development of individual model components**

153 The key hypothesis of our model is that storm properties can be represented using model compo-
154 nents that are functions of a small number of key local environmental variables. First, the genesis
155 model determines the rate at which weak vortices are formed throughout the domain, which are
156 then passed to the intensity and track models to determine the rest of the storms' life cycles.

157 *a. Genesis - Tropical Cyclone genesis index (TCGI)*

158 The essential element in the genesis model is the seeding rate. Previous studies have shown
159 that with only a few crucial environmental parameters, various TC genesis (potential) indices can
160 capture the location, frequency, and the seasonality of TC formation, including ENSO-induced
161 variability (Emanuel and Nolan 2004; Camargo et al. 2007b,a; Emanuel 2010; McGauley and
162 Nolan 2011; Tippett et al. 2011; Bruyère et al. 2012). Menkes et al. (2012) compare the existing
163 indices, and find that all have similar performance in genesis climatology. The Tropical Cyclone
164 Genesis Index (TCGI, developed by Tippett et al. 2011), however, has the least bias and the best
165 simulated seasonality. Thus, we calculate the seeding rate based on TCGI:

$$TCGI = \exp(b + b_{\eta}\eta_{850} + b_{RH}RH_{600} + b_{SST}SST_r + b_{SHRD}SHRD + \log(\cos\phi)). \quad (1)$$

166 The TCGI is the expected number of genesis events. η_{850} , RH_{600} , SST_r , $SHRD$ are the absolute
167 vorticity at 850 hPa, the relative humidity at 600 hPa, relative SST (SST relative to tropical mean
168 SST), and vertical shear between the 850- and 200- hPa levels. b is the intercept term and b_x s are
169 the coefficients corresponded to variable x . After fitting Eq. (1) with 32 years of inter-annually
170 varying data, we obtain a climatological relationship (b , and b_x) between observed genesis rate
171 and the predictors. We then apply such relationship to monthly data from 1981 to 2012 at spatial
172 resolution of 200 km to obtain monthly TCGI.

173 Using the TCGI for seeding rates, we select the grids and months where storms will form. For
174 each seed, a genesis location and date are then chosen randomly on a 1 km resolution within the
175 selected month. This seeding method allows the hazard model to form more than one vortex on the
176 same day at the same location, but this situation never occurs in our simulations. By construction,
177 the TCGI is always positive, and thus predicts a non-zero probability of storm formation globally
178 even in locations where no TC genesis events have been observed.

179 To evaluate the genesis model, we construct 40 realizations of 32-years simulations for the
180 period of 1981 to 2012. Globally, there are on average 95 storms per year and 11, 29, 23, 26, and
181 5 are in the ATL, WNP, ENP, SH, and IO, respectively. In the simulations, on average there are 94
182 storms per year with 8, 33, 18, 32, and 4 in each basin. The TCGI systematically underestimates
183 the genesis frequency in the ENP and ATL, and overestimates in the WNP and SH.

184 The spatial distributions of 32 years of genesis counts in observations (Fig. 1a) and based on
185 the TCGI (Fig. 1b) are in a good agreement. The TCGI has local maxima in approximately the
186 right locations, but with lower peak values and a smoother distribution. The observed highest TC
187 formation rate occurs in the ENP in observations, but is in the WNP in the TCGI. The simulated

188 distribution spreads further equatorward in the WNP and IO than in the observations. The for-
 189 mation rate in the central Pacific is higher than observed. The significance of these differences
 190 are shown in Fig. 1c using Z-score (Section 2). The negative genesis bias in ENP is statisti-
 191 cally significant with Z-score of 6 or higher. The biases in tropical Atlantic and southern Indian
 192 Ocean (negative) and in the southern Pacific and subtropical WNP (positive) are both considered
 193 significant but only with Z-score of 2–3. The Central Pacific bias and those at equators are not
 194 significant with Z-score around or smaller than 1. Additionally, Fig. 1c suggests that the TCGI
 195 errors are systematic, i.e., could be corrected.

196 *b. Track - Beta-advection model (BAM)*

197 After genesis, the track model moves the storm forward with an hourly time-step. Following
 198 Emanuel et al. (2006), we use a Beta Advection Model (BAM, ?). The BAM combines "beta
 199 drift" (Li and Wang 1994) with mean advection based on a linear combination of the large-scale
 200 low-level (850 hPa) and upper-level (250 hPa) winds:

$$\mathbf{V} = \alpha \mathbf{V}_{850} + (1 - \alpha) \mathbf{V}_{250} + \mathbf{V}_{\beta}, \quad (2)$$

201 \mathbf{V} is the vector of zonal (u) and meridional (v) wind time series at 850 and 200 hPa. α is a scalar
 202 weighting the winds at these two levels, and is set to 0.8 here. \mathbf{V}_{β} is the beta drift vector. The wind
 203 components are:

$$\begin{aligned} u_{250}(x, y, \tau, t) &= \bar{u}_{250}(x, y, \tau) + A_{11}F_1(t) \\ v_{250}(x, y, \tau, t) &= \bar{v}_{250}(x, y, \tau) + A_{21}F_1(t) + A_{22}F_2(t) \\ u_{850}(x, y, \tau, t) &= \bar{u}_{850}(x, y, \tau) + A_{31}F_1(t) + A_{32}F_2(t) + A_{33}F_3(t) \\ v_{850}(x, y, \tau, t) &= \bar{v}_{850}(x, y, \tau) + A_{41}F_1(t) + A_{42}F_2(t) + A_{43}F_3(t) + A_{44}F_4(t), \end{aligned} \quad (3)$$

204 in which \bar{u} and \bar{v} are daily resolution (τ) winds interpolated from monthly mean fields in a x and
 205 y grid. F_1 is a Fourier series variable with a random phase which represents a variability in winds
 206 for timescales(t) smaller than daily, an hour here:

$$F_1(t) \equiv \sqrt{\frac{2}{\sum_{n=1}^N n^{-3}}} \sum_{n=1}^N n^{-3/2} \sin[2\pi(nt/T + X_n)]. \quad (4)$$

207 In F , T is the lowest frequency (15 days) in the time series, N (15) is the total number of waves
 208 retained, and X_n is, for each n , a random number between 0 and 1. F_2 , F_3 , and F_4 have the same
 209 form as F_1 , but with different random phases, X_n . $A_{i,j}$ is the i th and j th coefficient in a lower
 210 triangular matrix \mathbf{A} that satisfies

$$\mathbf{A}^T \mathbf{A} = \text{COV}, \quad (5)$$

211 where COV is the covariance matrix of the flow components. \mathbf{A} is function of x, y , and τ .

212 The coefficient $n^{-3/2}$ in Eq. (4) is chosen to mimic the observed spectrum of geostrophic turbu-
 213 lence. The power spectrum of the kinetic energy of the synthetic winds from Eq. (4) falls close to
 214 the inverse cube of the frequency, and is steeper than that of the steering flow based on daily winds
 215 from reanalysis data (not shown). In short, Eq. (3) and (4) provide synthetic winds at 850 and 250
 216 hPa whose monthly means, variances, and covariances match those of reanalyses data.

217 Statistics of storm tracks are highly related to genesis location. The observed track density is
 218 roughly in phase with the observed genesis distribution (comparing Fig. 2a to Fig. 1a). In order
 219 to separate the BAM's performance from the genesis bias, we conduct 20 track realizations using
 220 the 32 years' observed genesis locations, using the simulated tracks with the same lifetimes as the
 221 best-track data.

222 Two experiments are conducted with different values for \mathbf{V}_β . In the first experiment, we set
 223 $\mathbf{V}_\beta = (0.0, 2.5)$ following Emanuel et al. (2006), that is, zero beta drift in the zonal direction and

224 2.5 ms^{-1} in the meridional direction. This setting is called “Ubeta0”. A recent study by Nakamura
 225 et al. (2017) shows a systematical north-northeast-ward track bias in Emanuel’s dataset in the
 226 WPC. Such bias might be related to the zero beta drift, which prevents westward moving tracks.
 227 Therefore, in the second experiment we choose V_β as a function of the cosine of latitude, with a
 228 maximum of 2.5 ms^{-1} . The cosine function is used because the β -drift changes with Coriolis force
 229 (Zhao et al. 2009). We call this second experiment “betaLat.”

230 The spatial distributions of the observed tracks and both experiments are in good agreement.
 231 This is primary because they have the same initial location. The spatial correlations between
 232 observations and Ubeta0 and betaLat are both very high (above 0.9). While there is no clear
 233 reason, based on these results alone, to view one as the better than the other, the fact that betaLat
 234 is more physics-based makes it more attractive, and we choose it here.

235 *c. Intensity - Autoregressive (AR) model*

236 The AR intensity model:

$$V_{t+12h} - V_t = L(V_t, V_{t-12h}, X_t, X_{t+12h}) + \varepsilon_{t+12h} \quad (6)$$

237 was described in our previous studies, Lee et al. (2015, 2016a). We refer readers to these two
 238 studies for details of the intensity model. Here we describe its general structure. V_t is the storm
 239 intensity at time t and X are environmental variables related to TC intensification. The deter-
 240 ministic component, $L(V_t, V_{t-12h}, X_t, X_{t+12h})$, has the form of a second-order vector autoregressive
 241 linear model with environmental variables as exogenous inputs. To predict intensity at $t+12h$, L
 242 includes storm information, V_t , $V_t - V_{t-12h}$, V_t^2 , and the storm translation speed. Three essential
 243 environmental variables, potential intensity (PI, Bister and Emanuel 2002; Camargo et al. 2007b),
 244 800–200 hPa deep layer mean vertical wind shear (SHR, Chen et al. 2006), 500–300hPa midlevel

245 relative humidity (midRH), are sufficient to reasonably simulate the storm intensity statistics (Lee
246 et al. 2015). PI enters L in the form of the difference between PI and initial storm intensity
247 $(PI - V_t)$, and its square and cubic forms: $(PI - V_t)^2$ and $(PI - V_t)^3$.

248 The stochastic forcing component (ε) accounts, in a statistically representative sense, for the
249 internal storm dynamics or other physical processes that do not depend explicitly on the environ-
250 ment. In other words, ε is the forecast error resulting from the linear assumption and the limited
251 variables included in L . Assuming that the forecast error is uncorrelated in time, i.e., white noise,
252 we randomly draw ε from the training period errors in conditioned on the initial intensity V_t . Lee
253 et al. (2016a) showed that including the white-noise stochastic term improves the simulated LMI
254 distribution as well as the spatial distribution of Cat3-5 storms. When a storm is close to land or
255 when it makes landfall, we switch the intensity model to the one that is fitted with an additional
256 parameter representing the surface type in L . Gray lines in Fig. 3a are the AR simulated LMI
257 distribution using the observed tracks and those in Fig. 3b are the landfall intensity distribution.
258 For LMI, the AR model captures the observed (black line) first peak but not the second small peak.
259 In the case of the landfall intensity, there is a small leftward shift representing a low bias in the
260 simulations.

261 **4. A TC hazard model**

262 The next step is to integrate all three components together to form a TC hazard model and to
263 evaluate model performance by its capability of simulating TC statistics. When all three compo-
264 nents are fully interactive, we refer to the solutions with the label, GTI, where ‘G’, ‘T’, and ‘I’
265 represent the genesis, track, and intensity respectively. For each synthetic storm, the initial inten-
266 sity is taken from the observed global distribution, not taking into account the basin-dependent
267 values of initial storm intensity (15-35 kt for the ATL and ENP, 15-30 for the other basins). The

268 dissipation is defined as the time when the intensity drops below 10 kt. We examine the storms'
269 evolution and only keep those which intensify and reach at least tropical storm (TS) strength (LMI
270 larger than 34 kt).

271 In GTI, only $70\pm 1\%$ of seeds become TS. This is because TCGI gives a non-negative chance
272 for storm formation globally, which can result in some initial seeds starting in very unfavorable
273 environments. Similarly, BAM can move the storm to an unfavorable environment since it only
274 knows the steering flow. Both situations lead to a frequency bias because TCGI is trained to
275 match the genesis of tropical storms (whose lifetime maximum intensity is at least 35 kt), not
276 the formation of the disturbed weather that can potentially become a tropical cyclone. In order
277 to maintain realistic global mean storm numbers, we revise the GTI simulations by seeding more
278 storms, factor of 1.4, globally than what the TCGI suggests. While the survival rate varies by
279 basin, we do not use a basin-dependent seeding rate. We will, however, apply a local frequency
280 adjustment when conducting hazard assessment (in Section 5).

281 After adjusting the survival rate, GTI generates synthetic storms whose climatology is in good
282 agreement with the observed one (Fig. 4). They both have more intense storms in the WNP and
283 less in the ATL, a westward followed by a north-eastward movement in the northern hemisphere,
284 and almost no storms in the southeastern Pacific and southern Atlantic. There are some differences
285 as well, such as more central Pacific storms and less pronounced equatorial gap in the simulations.
286 In addition to GTI, we designed two more experiments to isolate the influence of individual com-
287 ponents on the total estimated TC statistics: $\hat{G}TI$, and $\hat{G}\hat{T}I$. When ($\hat{\cdot}$) is used above these letters,
288 observational data are used instead of simulations. We construct 400 realizations of 32-year global
289 simulations (1981-2012) in each experiment (see Section 2 for details). In $\hat{G}TI$ and $\hat{G}\hat{T}I$, the ob-
290 served initial intensities are used for the corresponding formation locations.

291 *a. Genesis density and interannual variability*

292 $\hat{G}\hat{T}I$ and $\hat{G}TI$ genesis climatology (not shown) are similar to the observed one because best-track
293 genesis locations were used. Similarly, the spatial distribution of genesis location in GTI (Fig. 5a)
294 is close to $TCGI$ (Fig. 1b) in Section 3. This is because track and intensity models, while they
295 determine the survival of initial vortices, do not largely alter the genesis climatology. They do,
296 however, enhance the positive bias in central Pacific and WNP (comparing Fig. 1c and Fig. 5b),
297 which might be due to too many storms surviving in the central Pacific in GTI .

298 The interannual variability of storm frequency in individual basins is shown in Fig. 6. The
299 correlation coefficient for ATL hurricanes in GTI is 0.48, similar to Emanuel et al. (2008) while
300 with the new intensity model, it increases to 0.7 in Emanuel (2017). The correlation coefficient for
301 WNP, ENP, SH, and IO in GTI are 0.30, 0.36, 0.46, and -0.27, respectively. Menkes et al. (2012)
302 found that the existing genesis indices, including the $TCGI$, do not capture the full spectrum of
303 interannual variability in storm frequency well, although they are all able to simulate the impact
304 of ENSO. This feature is inherited in our model.

305 *b. Track and landfall frequency*

306 The track density plots from observations, $\hat{G}\hat{T}I$ and GTI are shown in Figs. 2a, 7a, and 7c. The
307 $\hat{G}\hat{T}I$ track density is the same as to observations and is not shown here. In both observations
308 and simulations, the highest value of the track density are in the WNP and ENP, followed by the
309 southern Indian Ocean and the western South Pacific. Both simulations show the typical observed
310 recurvature track pattern in the ATL. The relatively high track densities over northwestern Aus-
311 tralia and the Bay of Bengal, however, are missing in the simulations. A comparison between
312 model biases from $\hat{G}TI$ (Fig. 7b) and GTI (Fig. 7d) suggests that the negative frequency bias in
313 the ENP is due to the $TCGI$, consistent with the results from Section 3a. The positive frequency

314 bias in the central Pacific, which is also seen in the genesis Z -score in Fig. 5b, extends further
315 northwestward in Fig. 7d.

316 The regional landfall frequency in GTI simulation is in a good agreement with in observations
317 (Fig. 8). There is a low bias at the northern Indian Ocean (Fig. 8a) and Mexico to New England
318 (Fig. 8g) coastal regions, where the observed (black) frequency is constantly above the simulated
319 spread (red patches). The rank histograms (Section 2) also tilt towards high ranks in these regions
320 (Fig. 9). In these regions, we also see negative biases in the track density (Fig. 7d). In Taiwan
321 and the Philippines, there is a positive landfall frequency bias and the rank histogram distribution
322 tilts towards low ranks. The track density plot, however, shows a negative bias near Taiwan. This
323 inconsistency between biases in track density and landfall frequency occurs because the landfall
324 frequency is calculated at much finer spatial resolution (50 km) than is the track density (about
325 500 km). Thus, Taiwan covers only part of a large grid box in the track density plot. Another
326 possible reason is that landfall is related to the direction storm is moving. A low track frequency
327 does not necessary result in a low landfall occurrence if the number of westward moving tracks is
328 higher. The simulated landfall frequency is unbiased in the coastal regions from Vietnam to China.

329 *c. LMI and landfall intensity*

330 In Figure 3a, \hat{GTI} captures the first peak of the LMI distribution well, but misses the second
331 smaller peak due to an insufficient number of simulated RI storms (TCs that intensify, at least
332 once, more than 35 kt within 24 hours in their lifetimes), consistent with the results in Lee et al.
333 (2016a). PDFs of LMI from \hat{GTI} (light blue lines) and GTI (red lines), however, successfully
334 capture the observed second peak. This improvement has a simple explanation – the consistency
335 between track and intensity evolution. In \hat{GTI} , each synthetic storm ends when the observed
336 record ends, regardless of the storm's intensity at that time. As a result, some die while they are

337 intensifying, or still at or above TS level, and thus are artificially denied future opportunities to
338 undergo RI. Coupling the intensity model to the track model (in $\hat{G}TI$ and GTI) resolves this artifact
339 by giving each synthetic storm a self-consistent opportunity to undergo RI when the environment
340 permits. Thus, GTI and $\hat{G}TI$ generate numbers of RI storms close to those found in observations
341 (e.g., Fig. 10) and match the observed LMI distribution. The successful simulation of RI storms
342 shows that the stochastic forcing in the intensity model, as proposed in Lee et al. (2016a), is an
343 effective way to produce RI storms and gives further evidence that RI and the LMI distribution are
344 related. PDFs of the landfall intensity (Fig. 3b) in GTI and $\hat{G}TI$ are almost indistinguishable from
345 the observed one. Coupling between track and intensity model improves not only the simulation
346 of peak intensities, but the intensity evolution throughout storms' lifetimes as well, including at
347 landfall.

348 **5. Tropical cyclone hazard in the current climate**

349 In Section 4c and b, we discussed the performance of GTI in predicting TC landfall frequency
350 and intensity, respectively. When considering hazard, however, it is essential to use joint mea-
351 sures that contain both information about both of them. Thus, here we define TC hazard as the
352 probability of the landfall intensity exceeding a given threshold at a particular location. TC hazard
353 will be calculated based on the historical record, and synthetic storms from the the three simu-
354 lations, namely, $\hat{G}TI$, $\hat{G}TI$, and GTI . We will discuss TC hazard from both global and regional
355 perspectives.

356 *a. Global map of return period*

357 Figure 11 shows global maps of return period for hurricanes (Cat1+ storms) in observations and
358 simulations. At the coastal regions in the south western WNP (southeastern China, Taiwan, and the

359 Philippines), the observed return period of hurricanes is less than 10 years; it is close to 2-3 years
360 near Taiwan and the Philippines. Another distinct area with a low-return period (high hazard) is
361 the ENP. In the southern hemisphere, the 10-year return period occurs reach eastern Madagascar.
362 In the Northern Australia, Bay of Bengal and most US coastal regions, the return period for Cat1+
363 storms is on the order of decades. Because TCs are rare events, the ‘observed hazard’ does not
364 actually represent the true hazard, but is based on the length of the reliable historical observations.

365 Comparing the simulated return period maps of hurricane strength in Fig. 11 shows the ad-
366 vantages of using observed tracks and formation locations. Fig. 11b is much closer to Fig. 11a
367 (observations), than Figs. 11c (\hat{GTI}) and 11d (GTI) are. Some of the biases in the TC climatology
368 discussed earlier are reflected in the return period map. For example, GTI estimates a higher haz-
369 ard (shorter return period) in the central Pacific than do the observations. This difference is related
370 to the overestimation of storm activity in that area shown in Figs. 7b and 7d. In IO, the simulated
371 hazard is smaller than the observed one, due to the low frequency and the low intensity biases.
372 Despite these differences in detail, GTI captures the primary structure of high hazard regions for
373 hurricane strength storms.

374 The return period map of Cat4+ storms (Fig. 12) shows the advantages of calculating storm
375 evolution in a consistent environment, i.e., in GTI, for more rare events. \hat{GTI} underestimates the
376 Cat4+ storm hazard, especially in the WNP. GTI, on the other hand, reasonably capture the global
377 hazard of Cat4+, although the distribution is smoother compared to the observed map. This is
378 again because GTI is able to simulate sufficient numbers of RI storms.

379 *b. Regional return period*

380 To discuss TC hazard at regional scales, we select 13 sub-basin areas and calculate the return
381 period curves as a function of landfall intensity. The 13 chosen areas are the coastal regions of

382 Madagascar, Bay of Bengal, Vietnam, China, the Philippines, Taiwan, Japan, western Mexico,
383 Caribbean islands, Gulf of Mexico, eastern US, Pacific islands (Papua New Guinea and eastern
384 Indonesia) and northern Australia (Figs. 13 and 14). The observed return period curves, especially
385 those for the strongest landfall intensity thresholds, are limited by the available observations. In
386 simulations, the spread of simulated hazard increases with intensity because the low-intensity
387 landfall hazard is mostly controlled by the tracks. $\hat{G}\hat{T}I$ has only one set of tracks by construction
388 — i.e., the observed tracks — and therefore has almost no spread. In the $\hat{G}TI$ and GTI , there are
389 10 sets of tracks that contribute to the spread. At higher intensity thresholds, the intensity model
390 ensemble contributes to the spread of the hazard estimation.

391 Ideally, the historic return period curve falls within the range of the simulated curves, an indi-
392 cation of an unbiased model. Biases in the return period curves have at least two general sources:
393 landfall frequency (the location of the curves) and intensity (the shape of the curves). Model
394 curves shifted towards the right (left) with steeper (lower) slope can be interpreted as underestima-
395 tion (overestimation) of TC hazard. The observed return period curves (black) lay in the simulated
396 spreads of $\hat{G}\hat{T}I$ (gray patches) in most places. Although the observed tracks are used in $\hat{G}\hat{T}I$, there
397 are still a shifts towards the right in the simulated return period curves in the Bay of Bengal, Viet-
398 nam, Japan, western Mexico, indicating that some of the observed landfalling storms dissipate in
399 simulations before making landfall. In Australia and the Philippines, $\hat{G}\hat{T}I$ underestimates land-
400 fall intensities. Including the track model ($\hat{G}TI$, cyan patches) results in underestimations in most
401 places, except in the Pacific islands where $\hat{G}TI$ has more landfalls than the observations. This is
402 consistent with the equatorial bias discussed in Section 4.

403 Using the same environment conditions for genesis, track, and intensity (GTI , red patches) im-
404 proves the estimated return period curves. There is a small landfall frequency bias in the coastal
405 regions of China, Taiwan, Japan, and Caribbean islands. GTI underestimates the landfall fre-

406 quency in Madagascar, Vietnam, western Mexico, Gulf of Mexico, and eastern US. The bias is
407 largest in Gulf of Mexico, followed by Bay of Bengal. Furthermore, GTI results in too many
408 landfall events in the Pacific islands due to the equatorward bias in the southern Pacific.

409 In order to bias-correct the frequency locally, we shift the return period curve (GTI adjust, purple
410 patches) to match the observed return period at the lowest threshold, which is the threshold with
411 the most observations, and is more reliable than the higher intensity thresholds. After shifting
412 the simulated curves, the observed curves fall within the spread in the simulations for most of the
413 regions, staying within the 25 to 75 percentiles (darker purple patches), except in China and the
414 Pacific islands. In China, the observed return period curve for landfall intensities larger than 80 kt
415 is at the low edge of the simulated spread, i.e., the hazard is overestimated. The overestimation is
416 much more severe in the Pacific islands.

417 **6. Summary and discussion**

418 This study describes a new, environmentally forced tropical cyclone (TC) hazard model. It is
419 composed of three model components that, together, represent the complete storm lifetime: a
420 genesis model (TCGI), a beta-advection track model (BAM) and an auto-regressive (AR) intensity
421 model. The TCGI and BAM are developed following Tippett et al. (2011) and Emanuel et al.
422 (2006), respectively, while AR intensity model is from our previous work (Lee et al. 2016a). The
423 TCGI defines the spatial and temporal formation rate (i.e., the numbers of storms that should
424 form at a given location within a given period) using the observed climatological relationship
425 between storm formation and absolute vorticity, relative humidity, relative sea surface temperature,
426 and vertical shear (Section 3a). After the initial seeding, the BAM moves vortices following
427 the synthetic steering flow (Section 3b). The synthetic wind has the statistics of the monthly
428 averaged winds but also contains high-frequency perturbations calculated from the daily variance

429 and covariance. The intensity model predicts the storm's evolution using a deterministic multiple
430 linear regression plus a stochastic component (Section 3c). In the deterministic component of
431 the intensity model, the TC intensity change is a function of potential intensity, deep layer mean
432 vertical wind shear, midlevel relative humidity, and storm intensity persistence. The stochastic
433 component represents the physical processes that are not considered in the deterministic model and
434 is necessary in order for the intensity model to simulate the observed distribution of TC intensity.

435 The model captures many aspects of TC genesis, track, intensity, and landfall statistics, includ-
436 ing their density distributions, probability density function (PDF) of storms' lifetime maximum
437 intensity (LMI) and landfall intensity, as well as the landfall frequency. The model has a positive
438 frequency bias in the central Pacific and in the equatorial region. A particularly interesting result is
439 that it captures the observed LMI PDF, which has a main peak and a "shoulder" at higher intensi-
440 ties. This finding is different from our previous study, Lee et al. (2016a), in which the realizations
441 were conducted using the AR intensity model and observed tracks. The observed shoulder feature
442 in the global LMI PDF (the regional LMI PDF are bimodals) appears to be due to the separation in
443 two mono-modal PDFs, one from storms which undergo rapid intensification (RI, intensity change
444 larger than 35 kt per 24 hours) and the other one from those which do not (Lee et al. 2016b). While
445 the AR intensity model running along the observed tracks is able to simulate RI storms, it does not
446 generate as many RI storms as are found in observations. The reason for this underestimation is
447 that some of the synthetic storms end when the observed track ends regardless of their intensities,
448 which artificially reduces the probability of RI. Combining the AR intensity model and the BAM
449 track model resolves the inconsistency, and gives the synthetic storms opportunity to undergo RI
450 when the environment permits. Self-consistent tracks and intensities improve not only the LMI
451 distribution but the storms' lifetime intensities, and therefore also landfall intensities.

452 With the well-simulated TC climatology, the model can estimate regional TC hazard reasonably
453 well. However, it predicts more landfalls in the western North Pacific and Pacific islands, and
454 fewer landfalls in the northern Atlantic and Indian Ocean than in observations. These landfall
455 biases lead to biases in the estimated TC hazards. These landfall biases can be corrected during
456 the post-processing through a local frequency adjustment. The large positive hazard bias for the
457 Pacific islands, however, remains because the model generates too many strong landfalling storms
458 there. These and other biases in the TC hazard can be corrected to some extent, so that the TC
459 hazard model can generate estimates of the probability of landfall at a given intensity that are in
460 agreement with observations at shorter return periods, while also giving estimates at longer return
461 periods where such estimates cannot be directly generated from observations.

462 While the environmental parameters used here are obtained from reanalysis, they can poten-
463 tially be obtained instead from global climate model under various climate scenarios. However,
464 when assessing hazard in a changing climate, it may be appropriate to choose somewhat different
465 predictors. For example, Camargo et al. (2014) showed that using saturation deficit and potential
466 humidity allows for a better representation of the response to mean climate warming than using
467 relative humidity, although both indices have similar behavior in current climate. Parameters used
468 in the intensity model might need some adjustments as well. Preliminary results using one of the
469 CMIP5 models (not shown) suggest that the TC hazard model is able to produce reasonable TC
470 climatologies in both current and future climates. One of the challenging issues will be how to
471 make appropriate bias corrections in the required predictors obtained from different climate mod-
472 els. Application of our model in such a climate change context, forced by a range of global climate
473 models, will be presented in a future publication.

474 *Acknowledgments.* The research was supported by the the New York State Energy Research and
475 Development Authority under the research grant NYSERDA GG012621.

476 **References**

477 AIR WORLDWIDE, 2015: The AIR hurricane model: AIR Atlantic tropical cyclone model
478 V15.0.1 as implemented in Touchstone V3.0.0. Tech. rep., AIR WORLDWIDE.

479 Bister, M., and K. A. Emanuel, 2002: Low frequency variability of tropical cyclone potential
480 intensity 1. Interannual to interdecadal variability. *J. Geophys. Res.: Atmospheres*, **107**, 4801.

481 Bruyère, C. L., G. J. Holland, and E. Towler, 2012: Investigating the use of a genesis potential
482 index for tropical cyclones in the North Atlantic basin. *J. Climate*, **25**, 8611–8626.

483 Camargo, S. J., K. A. Emanuel, and A. H. Sobel, 2007a: Use of a genesis potential index to
484 diagnose ENSO effects on tropical cyclone genesis. *J. Climate*, **20**, 4819–4834.

485 Camargo, S. J., A. H. Sobel, A. G. Barnston, and K. A. Emanuel, 2007b: Tropical cyclone genesis
486 potential index in climate models. *Tellus*, **59**, 428–443.

487 Camargo, S. J., M. K. Tippett, A. H. Sobel, G. A. Vecchi, and M. Zhao, 2014: Testing the per-
488 formance of tropical cyclone genesis indices in future climates using the HiRAM model. *J.*
489 *Climate*, **27**, 9171–9196.

490 Chen, S. S., J. A. Knaff, and F. D. Marks, 2006: Effects of vertical wind shear and storm motion on
491 tropical cyclone rainfall asymmetries deduced from TRMM. *Mon. Wea. Rev.*, **134**, 3190–3208.

492 Chu, J.-H., C. R. Sampson, A. Lavine, and E. Fukada, 2002: *The Joint Typhoon Warning*
493 *Center tropical cyclone best-tracks, 1945-2000*. 22pp, Naval Research Laboratory Tech. rep.
494 NRL/MR/7540-02-16.

495 Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of
496 the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597.

497 ECMWF, 2013: Era interim, monthly mean of daily means. ECMWF, URL [http://apps.ecmwf.int/
498 datasets/data/interim-full-moda/levtype=pl/](http://apps.ecmwf.int/datasets/data/interim-full-moda/levtype=pl/).

499 Emanuel, K., 2010: Tropical cyclone activity downscaled from NOAA-CIRES reanalysis, 1908–
500 1958. *J. Adv. Model. Earth Syst.*, **2**, doi:10.3894/JAMES.2010.2.1.

501 Emanuel, K., 2015: Effect of upper-ocean evolution on projected trends in tropical cyclone activ-
502 ity. *J. Climate*, **28**, 8165–8170.

503 Emanuel, K., 2017: A fast intensity simulator for tropical cyclone risk analysis. *Natural Hazards*,
504 **88**, 779–796.

505 Emanuel, K., C. DesAutels, C. Holloway, and R. Korty, 2004: Environmental control of tropical
506 cyclone intensity. *J. Atmos. Sci.*, **61**, 843–858.

507 Emanuel, K., R. Sundararajan, and J. Williams, 2008: Hurricanes and global warming: Results
508 from downscaling IPCC AR4 simulations. *Bull. Amer. Meteor. Soc.*, **89**, 347–367.

509 Emanuel, K. A., 2013: Downscaling CMIP5 climate models shows increased tropical cyclone
510 activity over the 21st century. *PNAS*, **110**, 12 219–12 224.

511 Emanuel, K. A., and D. S. Nolan, 2004: Tropical cyclone activity and global climate. *Bull. Amer.*
512 *Meteor. Soc.*, **85**, 666–667.

513 Emanuel, K. A., S. Ravela, E. Vivant, and C. Risi, 2006: A statistical deterministic approach to
514 hurricane risk assessment. *Bull. Amer. Meteor. Soc.*, **87**, 299–314.

515 Estrada, F., W. J. W. Botzen, and R. S. J. Tol, 2015: Economic losses from US hurricanes consistent
516 with an influence from climate change. *Nature Geosci*, **8**, 880–884.

517 Geiger, T., K. Frieler, and A. Levermann, 2016: High-income does not protect against hurricane
518 losses. *Environ. Res. Lett.*, **11**, 084012.

519 Hall, T., and S. Jewson, 2007: Statistical modelling of North Atlantic tropical cyclone tracks.
520 *Tellus*, **59**, 486–498.

521 JTWC, 2014: Tropical cyclone best track data site. Joint Typhoon Warning Center, URL http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/best_tracks.
522

523 Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation
524 of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592.

525 Lee, C.-Y., M. K. Tippett, S. J. Camargo, and A. H. Sobel, 2015: Probabilistic multiple linear
526 regression modeling for tropical cyclone intensity. *Mon. Wea. Rev.*, **143**, 933–954.

527 Lee, C.-Y., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2016a: Autoregressive modeling for
528 tropical cyclone intensity climatology. *J. Climate*, **29**, 7815–7830.

529 Lee, C.-Y., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2016b: Rapid intensification and the
530 bimodal distribution of tropical cyclone intensity. *Nat. Commun.*, **7**, 10625.

531 Li, X., and B. Wang, 1994: Barotropic dynamics of the beta gyres and beta drift. *J. Atmos. Sci.*,
532 **51**, 746–756.

533 Lin, N., K. Emanuel, M. Oppenheimer, and E. Vanmarcke, 2012: Physically based assessment of
534 hurricane surge threat under climate change. *Nature Clim. Change*, **2**, 462–467.

535 Lin, N., R. Jing, Y. Wang, E. Yonekura, J. Fan, and L. Xue, 2017: A statistical investigation of the
536 dependence of tropical cyclone intensity change on the surrounding environment. *Mon. Wea.*
537 *Rev.*, **145**, 2813–2831.

538 Marks, D. G., 1992: The beta and advection model for hurricane track forecasting. Tech. rep.,
539 Natl. Meteor. Center, Camp Springs, Maryland.

540 McGauley, M. G., and D. S. Nolan, 2011: Measuring environmental favorability for tropical cy-
541 clogenesis by statistical analysis of threshold parameters. *J. Climate*, **24**, 5968–5997.

542 Menkes, C. E., M. Lengaigne, and P. Marchesiello, 2012: Comparison of tropical cyclogenesis
543 indices on seasonal to interannual timescales. *Clim Dyn*, **38**, 301–321.

544 Nakamura, J., and Coauthors, 2017: Western North Pacific tropical cyclone model tracks in present
545 and future climates. *J. Geophys. Res.: Atmospheres*, early online, doi: 10.1002/2017JD02700.

546 NHC, 2013: Best track data (HRUDAT2). URL [http://www.nhc.noaa.gov/data/
547 #hurdat](http://www.nhc.noaa.gov/data/#hurdat).

548 Tippett, M., S. J. Camargo, and A. H. Sobel, 2011: A Poisson regression index for tropical cyclone
549 genesis and the role of large-scale vorticity in genesis. *J. Climate*, **21**, 2335–2357.

550 Tolwinski-Ward, S. E., 2015: Uncertainty quantification for a climatology of the frequency and
551 spatial distribution of North Atlantic tropical cyclone landfalls. *J. Adv. Model. Earth Syst.*, **7**,
552 305–319.

553 Yonekura, E., and T. M. Hall, 2011: A statistical model of tropical cyclone tracks in the western
554 north Pacific with ENSO-dependent cyclogenesis. *J. Appl. Meteor. Climatol.*, **50** (8), 1725–
555 1739.

- 556 Yonekura, E., and T. M. Hall, 2014: ENSO effect on East Asian tropical cyclone landfall via
557 changes in tracks and genesis in a statistical model. *J. Appl. Meteor. Climatol.*, **53**, 406–420.
- 558 Zhao, H., L. Wu, and W. Zhou, 2009: Observational relationship of climatologic beta drift with
559 large-scale environmental flows. *Geophys. Res. Lett.*, **36**, L18 809.

560 **LIST OF FIGURES**

561 **Fig. 1.** Number of TC genesis per $5^\circ \times 5^\circ$ box from 1981 to 2012 (a) from observations, and (b)
562 averaged from 40 TCGI simulations. (c) Z-score of TCGI simulations. Z-score lower than
563 1 is insignificant and is not shown here. Note that scale in (a) and (b) are logarithmic, and is
564 linear in (c). 30

565 **Fig. 2.** Track counts every $5^\circ \times 5^\circ$ box from 1981 to 2012 from (a) observations, (b) averaged from
566 20 BAM simulations with zero zonal beta component (Ubeta0), and (c) averaged from 20
567 BAM simulations with latitude-dependent beta drift (betaLat). The color scale is logarithmic.
568 In (b) and (c) the storms' genesis locations and lifetime are from observations. 31

569 **Fig. 3.** (a) LMI from 1981-2012 from observations (black), $\hat{G}\hat{T}I$ (gray), $\hat{G}T\hat{I}$ (cyan), GTI (red). (b)
570 Similar to (a) but for landfall intensity distribution. Each of the experiments contains 400
571 realizations. 32

572 **Fig. 4.** (a) 2000-2012 historical tracks color-coded by intensity. (b) Similar to (a) but from a ran-
573 domly selected member (out of 400 realizations) from GTI 33

574 **Fig. 5.** (a) Number of TC genesis per $5^\circ \times 5^\circ$ box averaged over 400 GTI simulations. (b) Z-score
575 of GTI simulations. Z-score lower than 1 is insignificant and is not plotted. The color scale
576 is logarithmic in (a) and linear in (b). 34

577 **Fig. 6.** Interannual variability of storm genesis in ATL (red), WNP (blue), ENP (green), SH (pur-
578 ple), IO (yellow). The observed time-series are in thick solid lines while the GTI simulated
579 ones are in thin solid lines with the thick dashed lines representing the ensemble mean. Data
580 are normalized by the corresponded mean and standard deviation and the black lines are
581 the reference lines (i.e., zero). The interval between two black horizontal lines is four stan-
582 dard deviations. The correlation coefficient between observations and ensemble means in
583 individual basins are given on the top of the figures. 35

584 **Fig. 7.** (a) Simulated 1981-2012 TC track counts per $5^\circ \times 5^\circ$ box from 400 ensemble mean from
585 $\hat{G}\hat{T}I$. (b) Z-score of the $\hat{G}\hat{T}I$ simulations. (c) Similar to (a) but from GTI. (d) Z-score of the
586 GTI simulations. The scales are logarithmic in (a) and (c) and linear in (b) and (d). 36

587 **Fig. 8.** Observed (black line) and GTI simulated (red shading) landfall frequencies in number of
588 occurrences at every 50 km along the coastline of (a) Northern Indian Ocean, (b) Vietnam
589 to China, (c) the Philippines, (d) Taiwan, (e) Japan, (f) Eastern Pacific, (g) Mexico to New
590 England, (h) northern Australia. The simulated landfall frequencies are shown as 0, 25, 75,
591 100 percentile based on the 400 realizations. X-axis in each panel matches with colors along
592 the corresponding coastline in the background map, starting from 'X' symbol. The color is
593 lighter with increasing distance. 37

594 **Fig. 9.** The normalized rank histogram from landfall frequencies from Fig. 8. 38

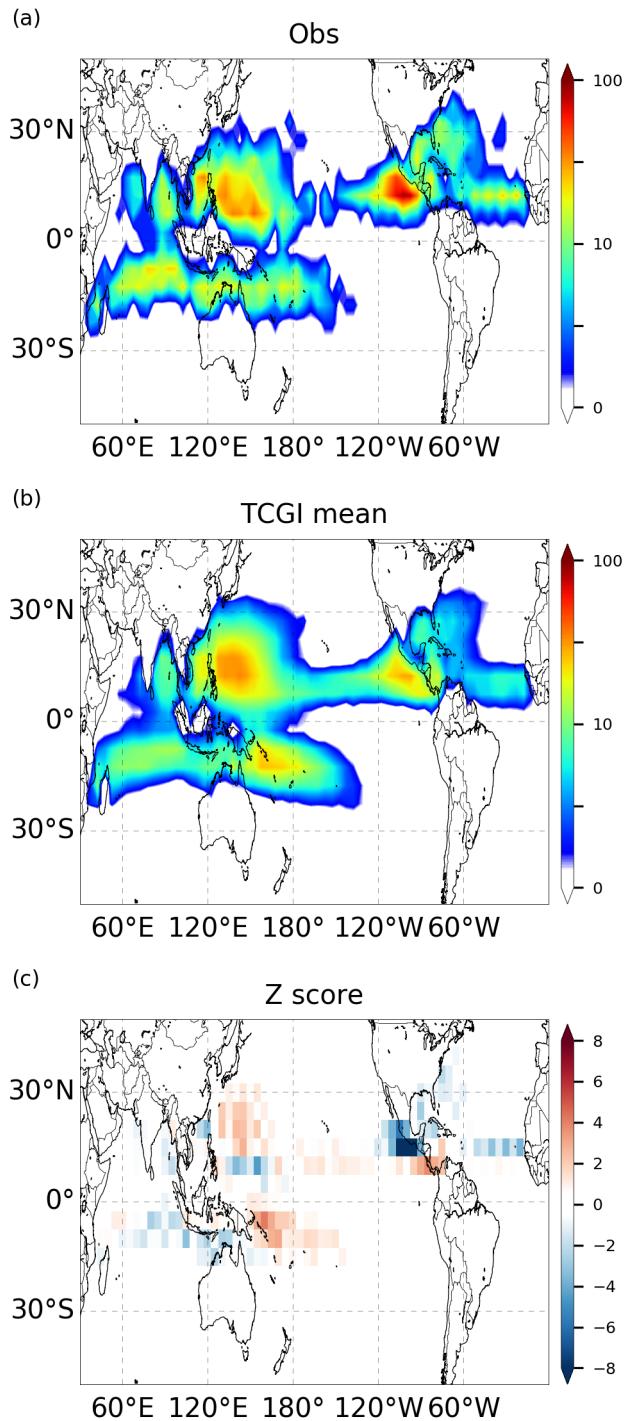
595 **Fig. 10.** PDF of LMI from 1981-2012 global historical record (black) and from 400 GTI realizations
596 (gray). Blue and red lines are PDFs using subsets of non-RI and RI storms from observa-
597 tions. Same are the cyan and pink lines but from simulations. 39

598 **Fig. 11.** Return period map for storms exceeding Category 1 hurricane strength from (a) 1981-2012
599 observations, and 12800-yr simulations from (b) $\hat{G}\hat{T}I$, (c) $\hat{G}T\hat{I}$, and (d) GTI. Data are cal-
600 culated in $2^\circ \times 2^\circ$, and a Gaussian smoothing is applied with length scale of 3 grid points.
601 40

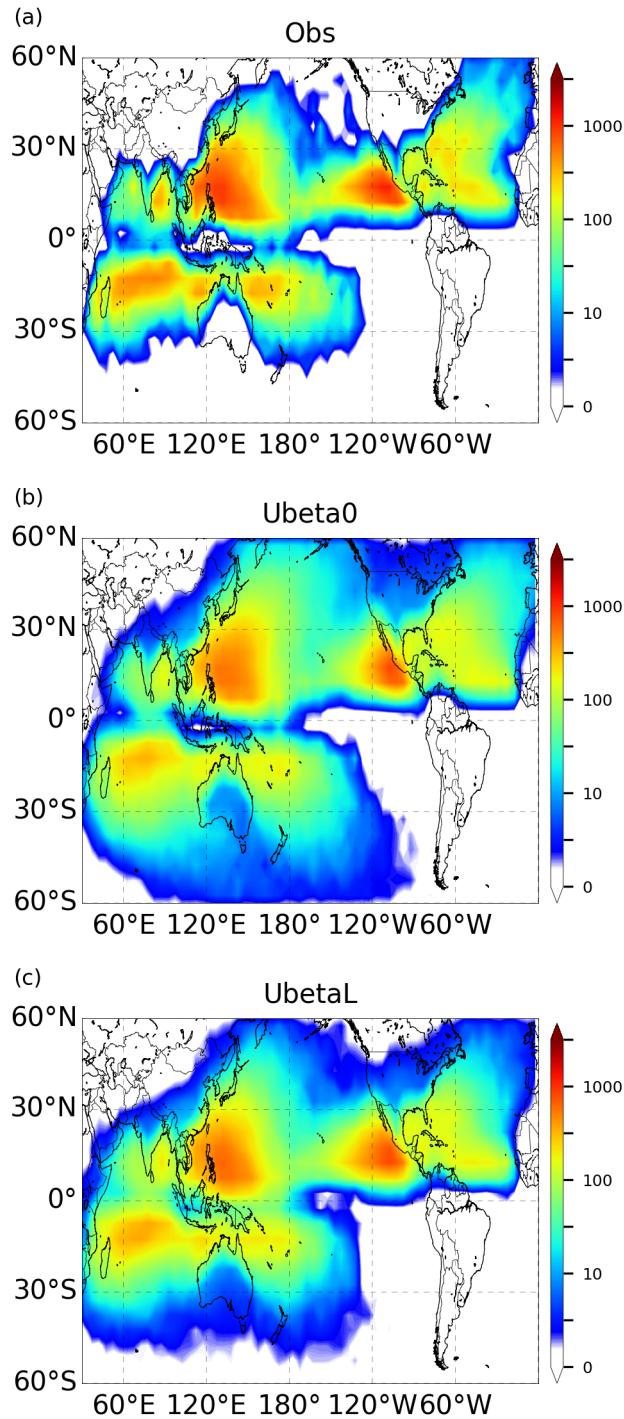
602 **Fig. 12.** Similar to Fig. 11 but for storms exceeding Category 4 hurricane strength. 41

603 **Fig. 13.** Return period curve of landfall intensity in (a) Madagascar,(b) Bay of Bengal, (c) Vietnam,
604 (d) China, (e) the Philippines, (f) Taiwan, and (g) Japan from observations (black), and 400
605 ensemble members from simulations (from left to right): $\hat{G}\hat{T}I$ (gray shading), $\hat{G}T\hat{I}$ (cyan
606 shading), GTI (red shading) and regional frequency-adjusted GTI (purple shading). The
607 solid lines are return period curves calculated using all 12800 year simulations. 42

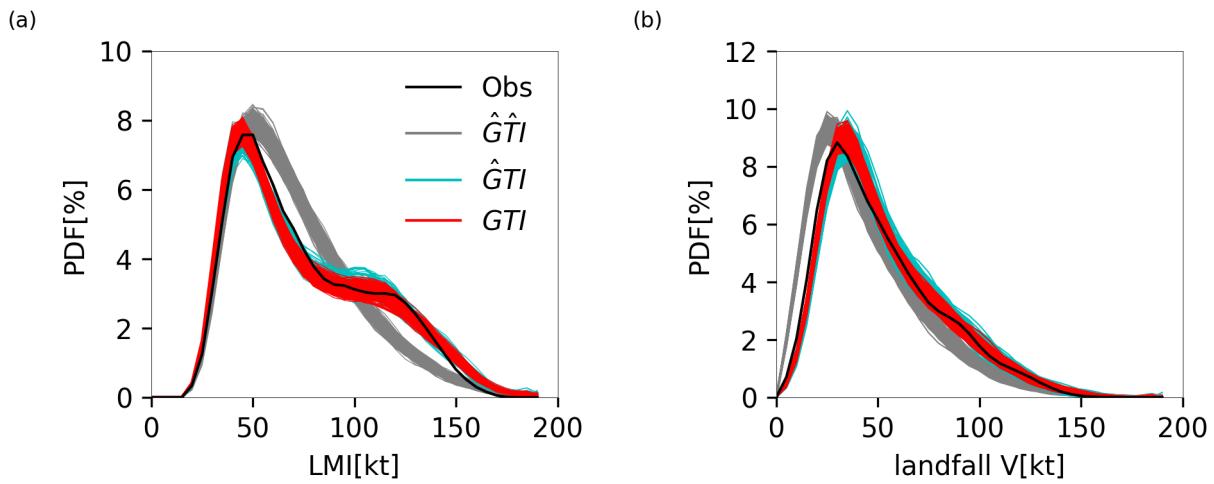
608 **Fig. 14.** Continuation of Fig. 13. (h) western Mexico, (i) Caribbean islands, (j) Gulf of Mexico, (k)
609 Eastern US, (l) Pacific islands, and (m) Australia. 43



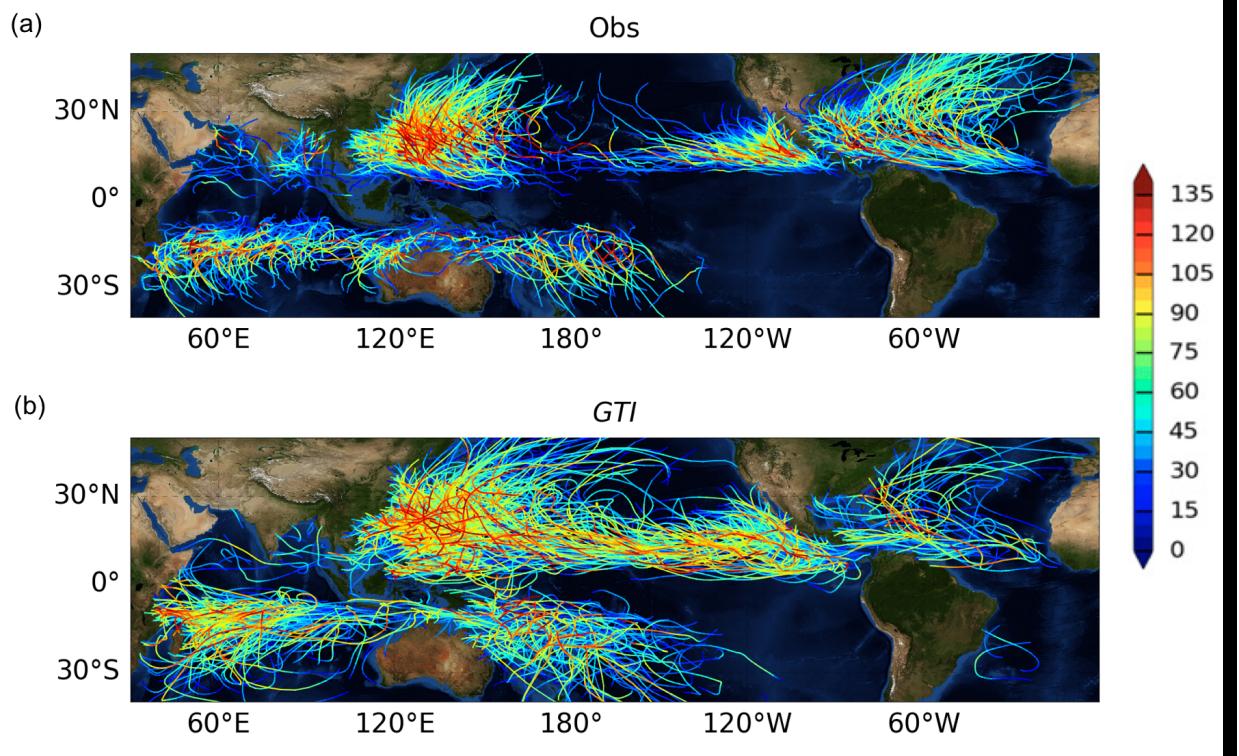
610 FIG. 1. Number of TC genesis per $5^\circ \times 5^\circ$ box from 1981 to 2012 (a) from observations, and (b) averaged
 611 from 40 TCGI simulations. (c) Z-score of TCGI simulations. Z-score lower than 1 is insignificant and is not
 612 shown here. Note that scale in (a) and (b) are logarithmic, and is linear in (c).



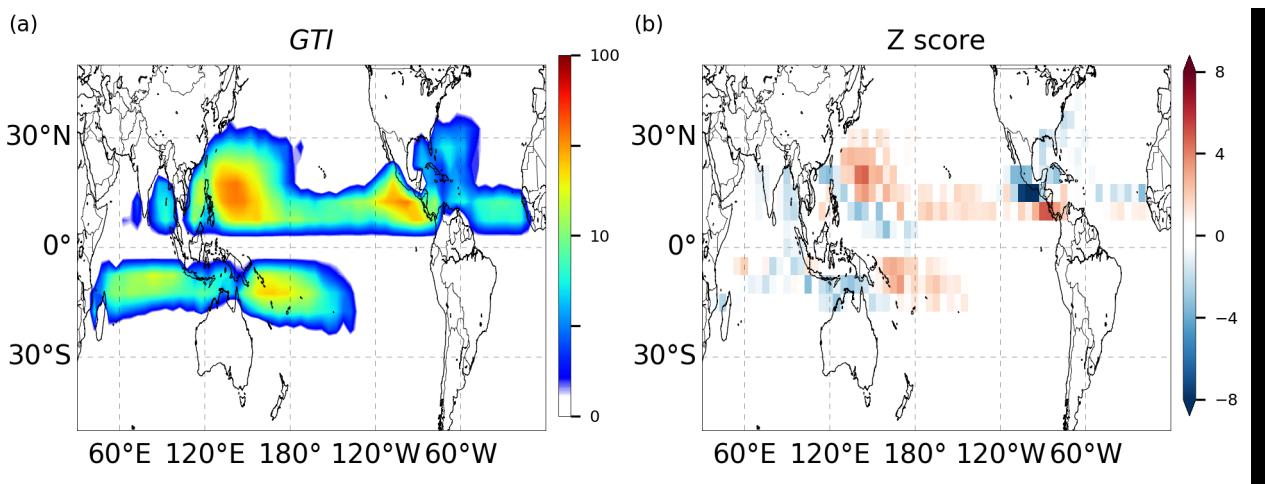
613 FIG. 2. Track counts every $5^\circ \times 5^\circ$ box from 1981 to 2012 from (a) observations, (b) averaged from 20 BAM
 614 simulations with zero zonal beta component (Ubeta0), and (c) averaged from 20 BAM simulations with latitude-
 615 dependent beta drift (betaLat). The color scale is logarithmic. In (b) and (c) the storms' genesis locations and
 616 lifetime are from observations.



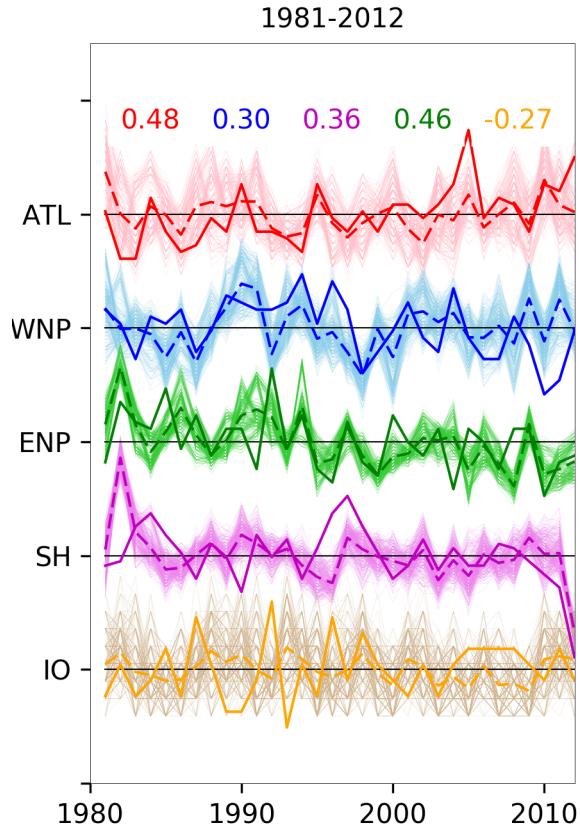
617 FIG. 3. (a) LMI from 1981-2012 from observations (black), \hat{G}^TI (gray), $\hat{G}TI$ (cyan), GTI (red). (b) Similar to
 618 (a) but for landfall intensity distribution. Each of the experiments contains 400 realizations.



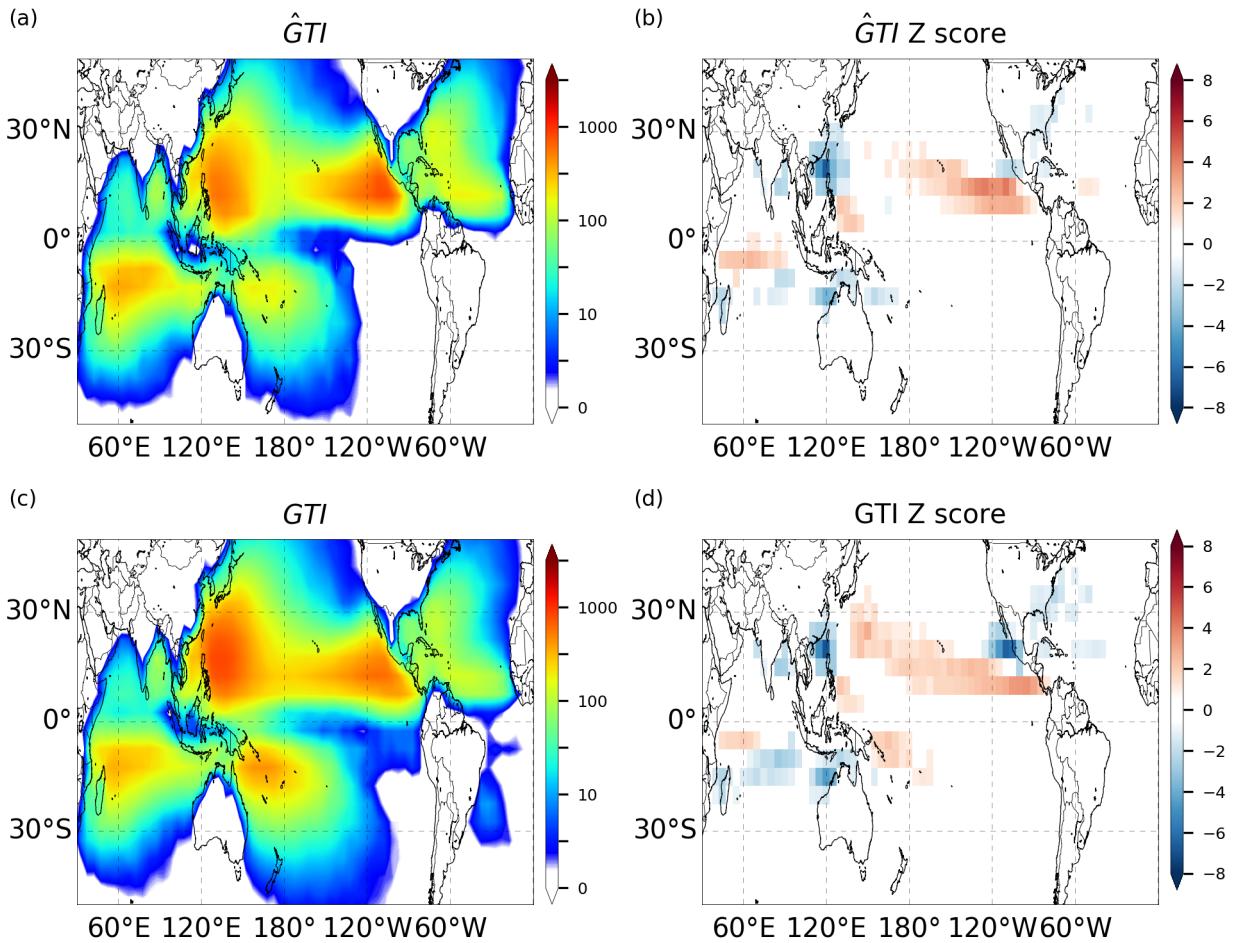
619 FIG. 4. (a) 2000-2012 historical tracks color-coded by intensity. (b) Similar to (a) but from a randomly
 620 selected member (out of 400 realizations) from GTI



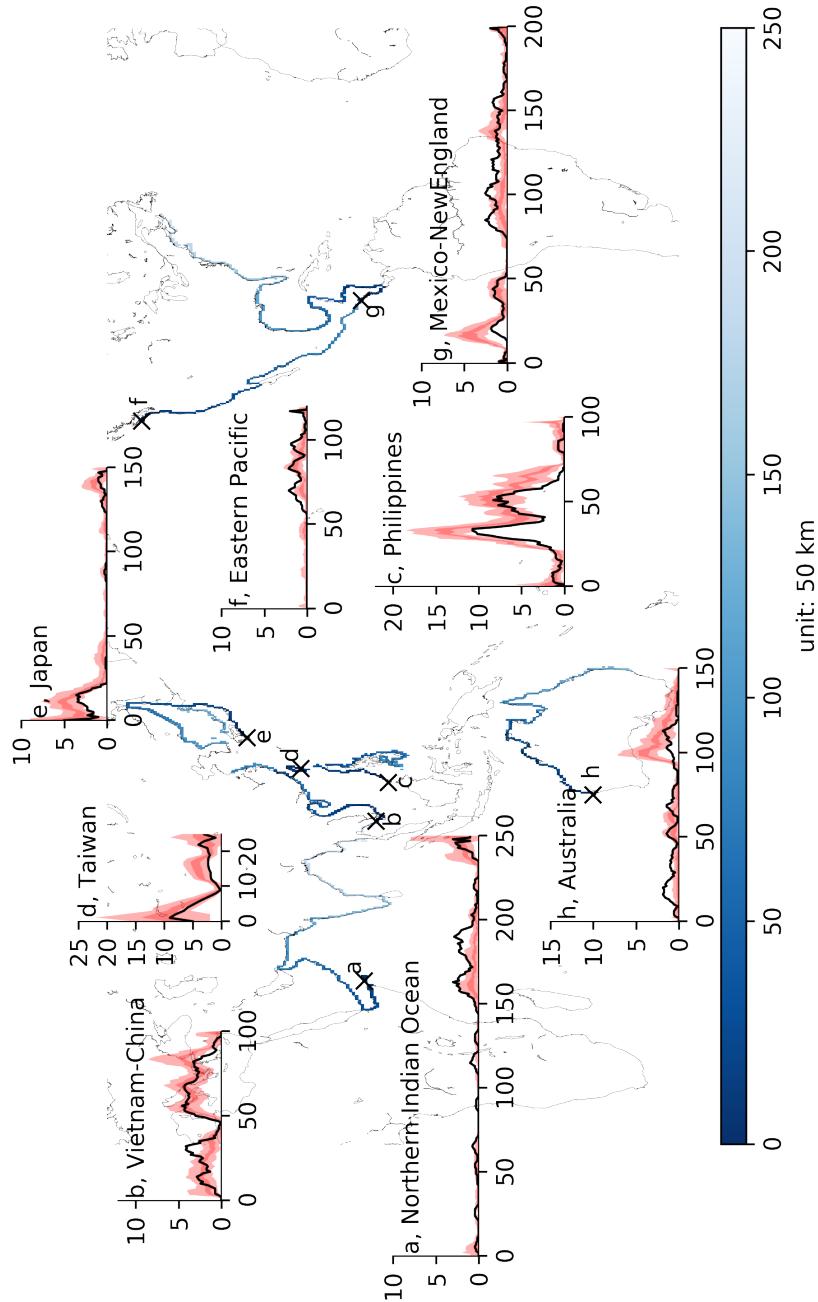
621 FIG. 5. (a) Number of TC genesis per $5^\circ \times 5^\circ$ box averaged over 400 GTI simulations. (b) Z-score of GTI
 622 simulations. Z-score lower than 1 is insignificant and is not plotted. The color scale is logarithmic in (a) and
 623 linear in (b).



624 FIG. 6. Interannual variability of storm genesis in ATL (red), WNP (blue), ENP (green), SH (purple), IO
 625 (yellow). The observed time-series are in thick solid lines while the GTI simulated ones are in thin solid lines
 626 with the thick dashed lines representing the ensemble mean. Data are normalized by the corresponded mean
 627 and standard deviation and the black lines are the reference lines (i.e., zero). The interval between two black
 628 horizontal lines is four standard deviations. The correlation coefficient between observations and ensemble
 629 means in individual basins are given on the top of the figures.



630 FIG. 7. (a) Simulated 1981-2012 TC track counts per $5^\circ \times 5^\circ$ box from 400 ensemble mean from \hat{GTI} . (b)
 631 Z-score of the \hat{GTI} simulations. (c) Similar to (a) but from GTI . (d) Z-score of the GTI simulations. The scales
 632 are logarithmic in (a) and (c) and linear in (b) and (d).



633 FIG. 8. Observed (black line) and GTI simulated (red shading) landfall frequencies in number of occurrences
 634 at every 50 km along the coastline of (a) Northern Indian Ocean, (b) Vietnam to China, (c) the Philippines,
 635 (d) Taiwan, (e) Japan, (f) Eastern Pacific, (g) Mexico to New England, (h) northern Australia. The simulated
 636 landfall frequencies are shown as 0, 25, 75, 100 percentile based on the 400 realizations. X-axis in each panel
 637 matches with colors along the corresponding coastline in the background map, starting from 'X' symbol. The
 638 color is lighter with increasing distance.

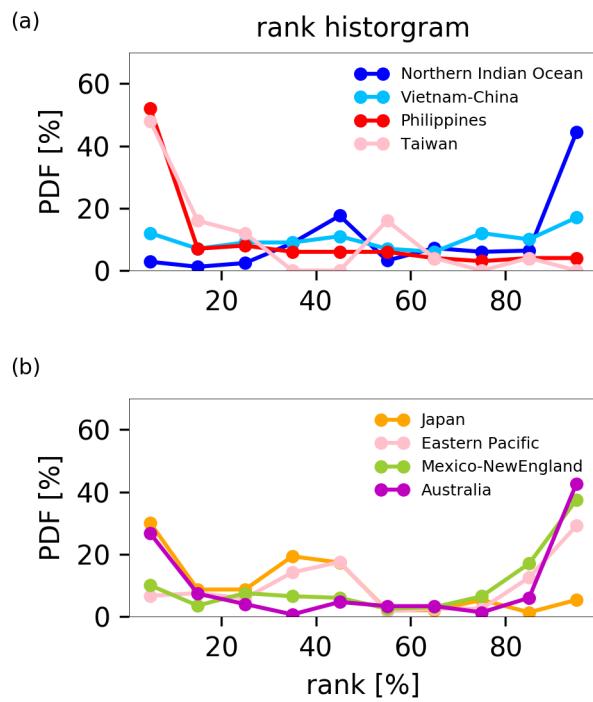
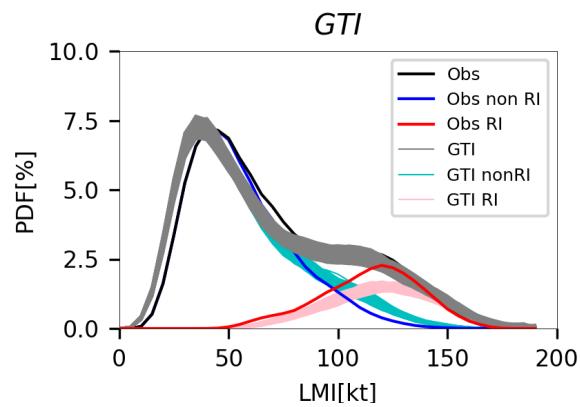
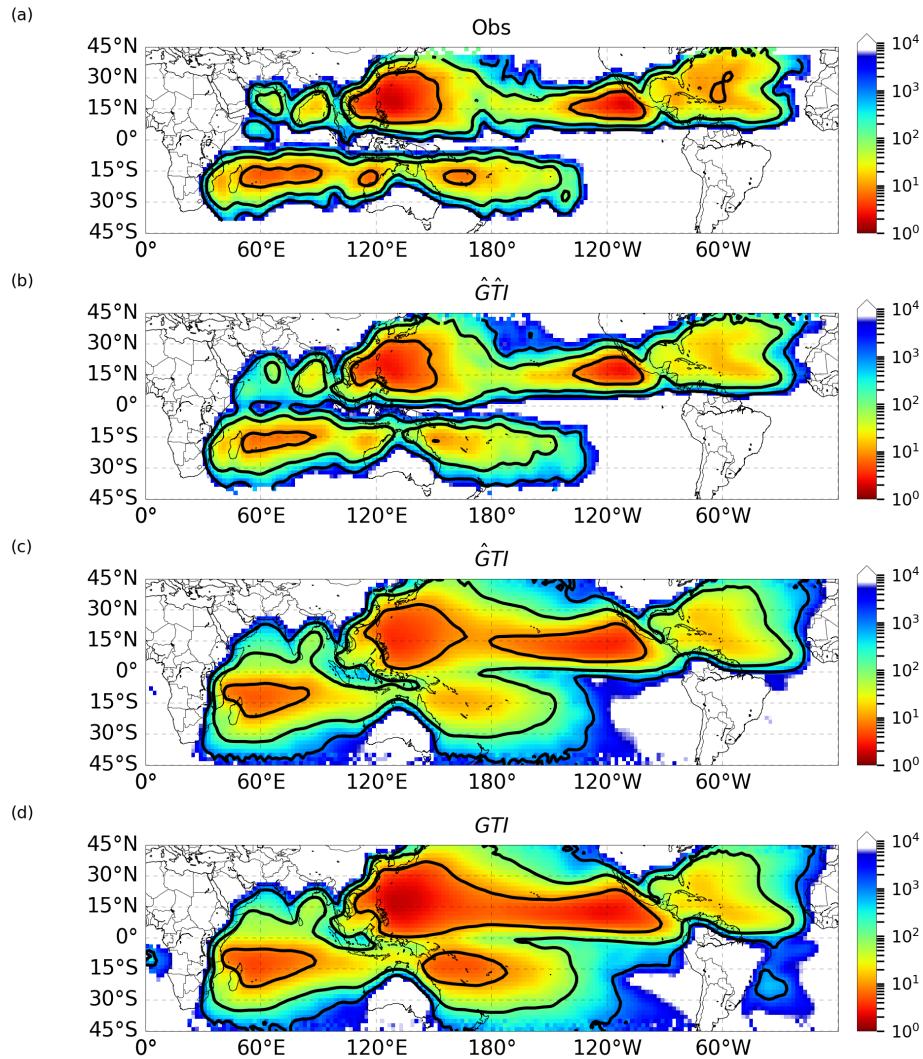


FIG. 9. The normalized rank histogram from landfall frequencies from Fig. 8.



639 FIG. 10. PDF of LMI from 1981-2012 global historical record (black) and from 400 GTI realizations (gray).
 640 Blue and red lines are PDFs using subsets of non-RI and RI storms from observations. Same are the cyan and
 641 pink lines but from simulations.



642 FIG. 11. Return period map for storms exceeding Category 1 hurricane strength from (a) 1981-2012 obser-
 643 vations, and 12800-yr simulations from (b) $\hat{G}TI$, (c) $\hat{G}TI$, and (d) GTI . Data are calculated in $2^\circ \times 2^\circ$, and a
 644 Gaussian smoothing is applied with length scale of 3 grid points.

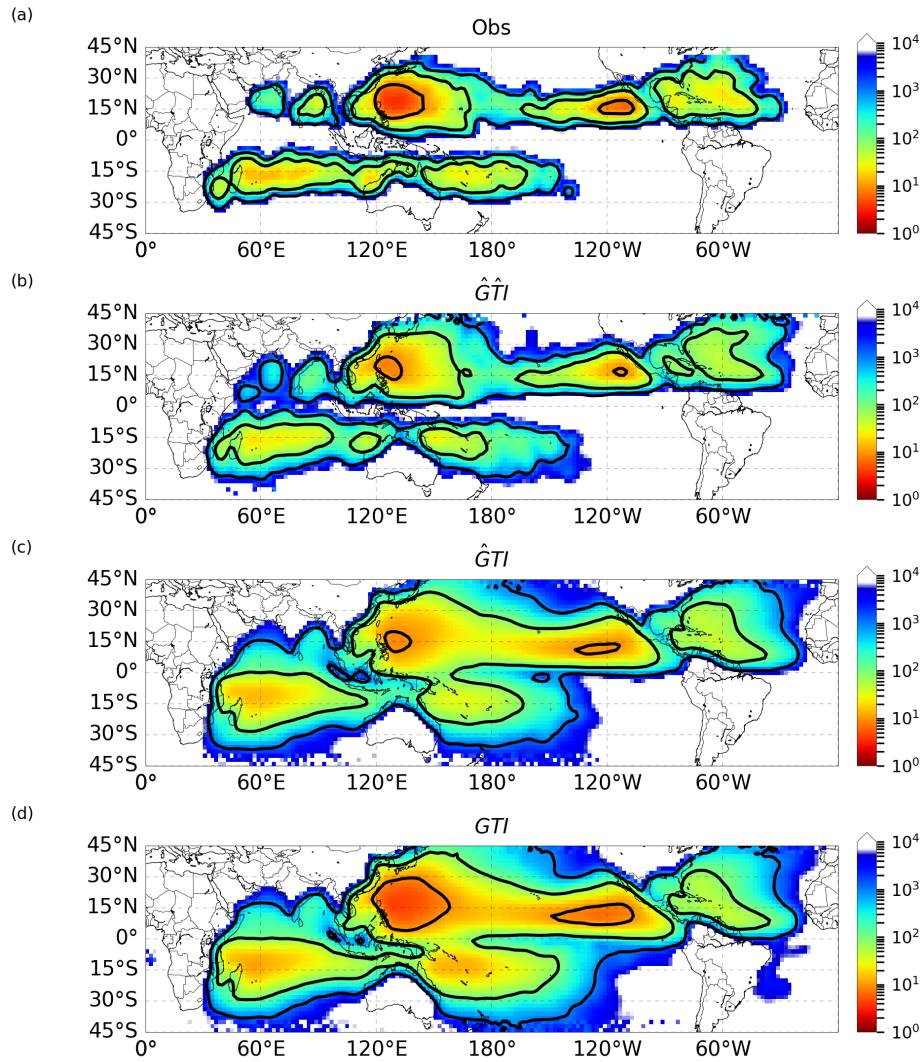
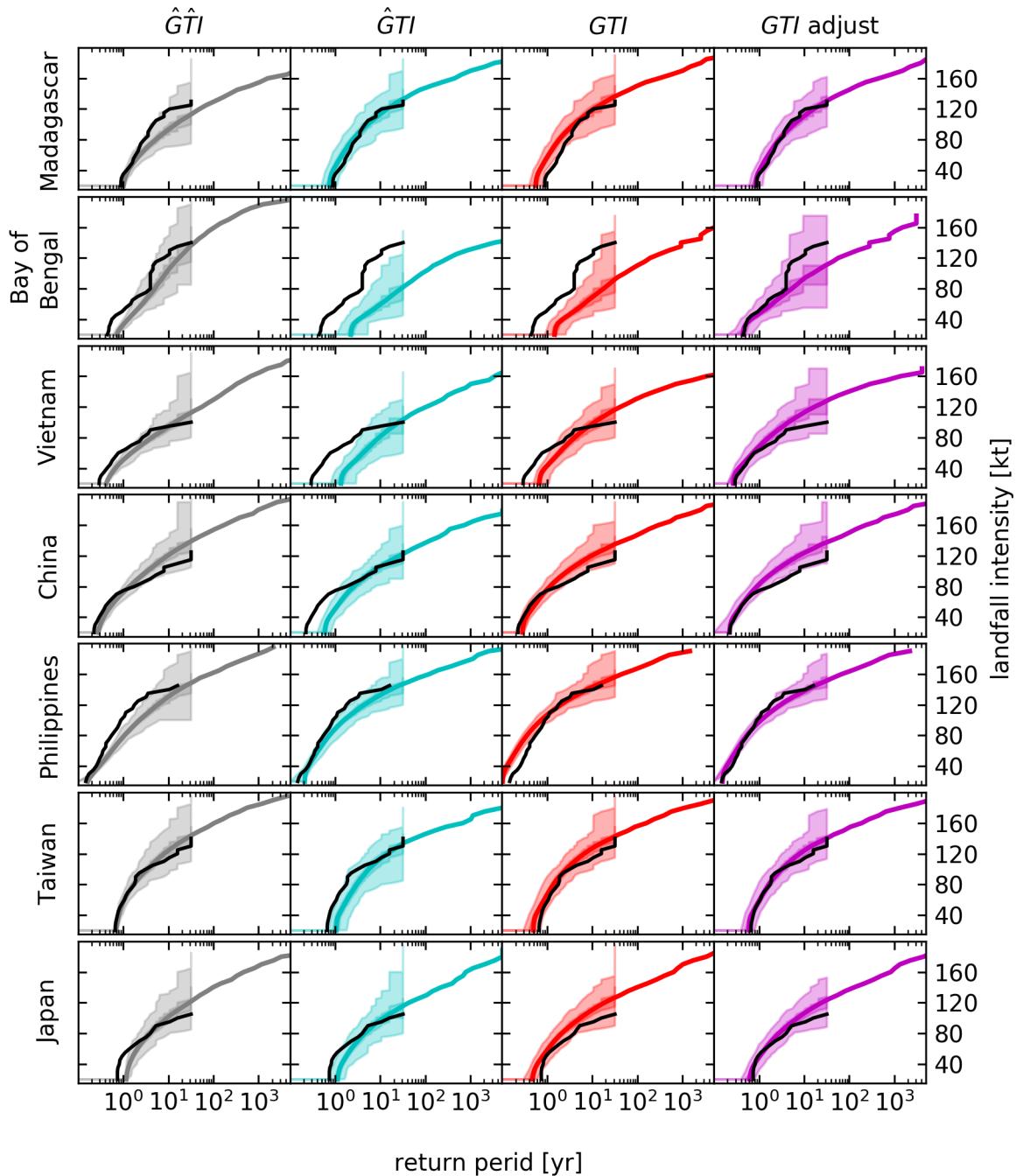
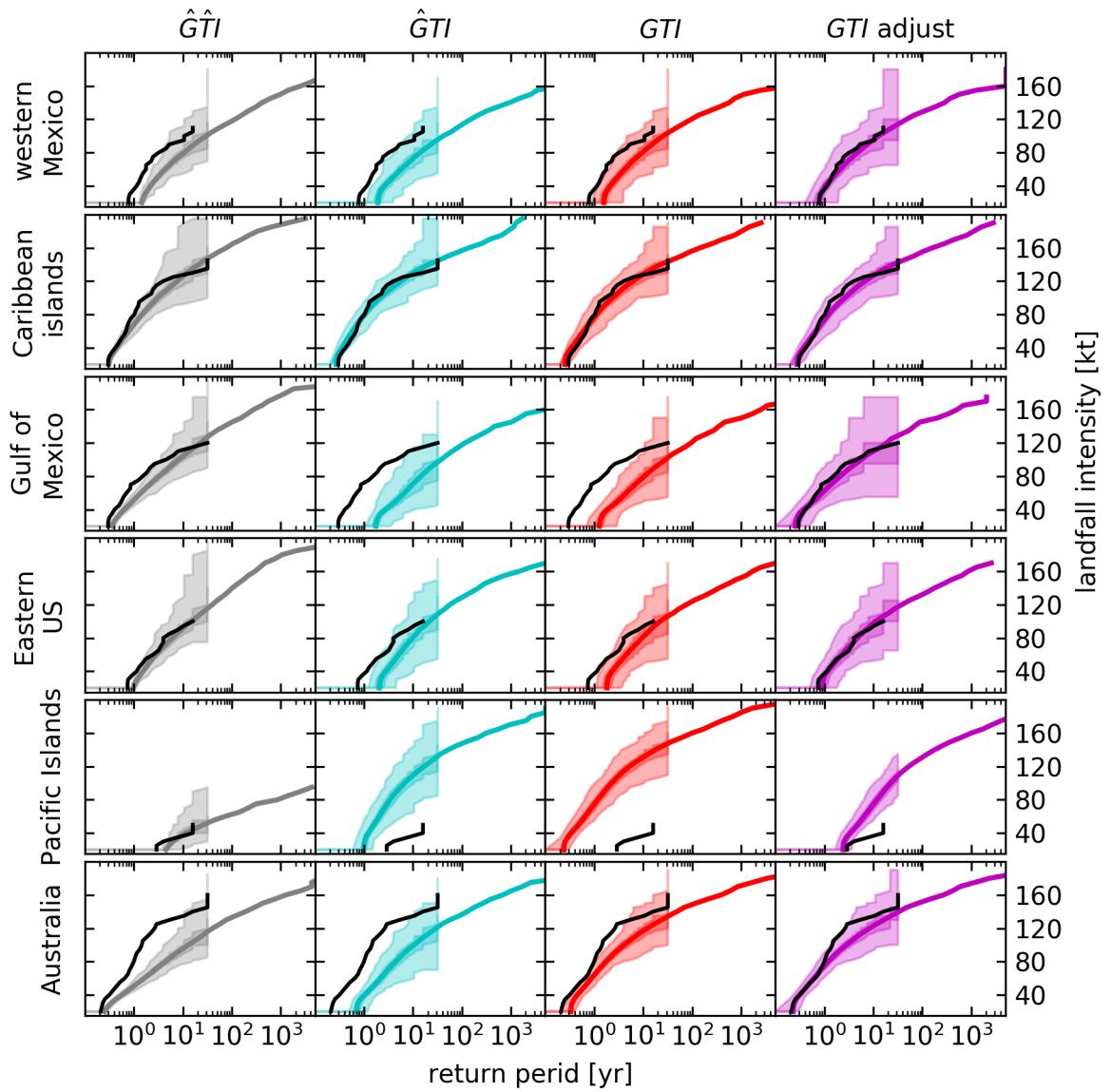


FIG. 12. Similar to Fig. 11 but for storms exceeding Category 4 hurricane strength.



645 FIG. 13. Return period curve of landfall intensity in (a) Madagascar, (b) Bay of Bengal, (c) Vietnam, (d)
 646 China, (e) the Philippines, (f) Taiwan, and (g) Japan from observations (black), and 400 ensemble members
 647 from simulations (from left to right): $\hat{G}TI$ (gray shading), $\hat{G}TI$ (cyan shading), GTI (red shading) and regional
 648 frequency-adjusted GTI (purple shading). The solid lines are return period curves calculated using all 12800
 649 year simulations.



650 FIG. 14. Continuation of Fig. 13. (h) western Mexico, (i) Caribbean islands, (j) Gulf of Mexico, (k) Eastern
 651 US, (l) Pacific islands, and (m) Australia.