Key Points:

- Changes in the Atlantic tropical cyclone (TC) risk are uncertain due to epistemic uncertainty in the projected annual frequency under global warming
- Likelihood analysis shows that observations are more consistent with simulations with upward frequency projections than those without
- It is more likely that the risk of increased TC impacts in Atlantic is increasing than that it is decreasing, though not by a large margin

Correspondence to:

C.-Y. Lee,
cl3225@columbia.edu

Author Contributions:

Conceptualization: Chia-Ying Lee, Adam H. Sobel
Data curation: Michael Wehner, Hiroyuki Murakami
Formal analysis: Chia-Ying Lee
Funding acquisition: Marc Wüest
Investigation: Chia-Ying Lee, Adam H. Sobel, Michael K. Tippett
Methodology: Chia-Ying Lee, Michael K. Tippett, Suzana J. Camargo
Project Administration: Marc Wüest
Software: Chia-Ying Lee
Validation: Chia-Ying Lee, Suzana J. Camargo
Visualization: Chia-Ying Lee

Climate Change Signal in Atlantic Tropical Cyclones Today and Near Future

Chia-Ying Lee[1] , Adam H. Sobel[1,2] , Michael K. Tippett[2] , Suzana J. Camargo[1] , Marc Wüest[3], Michael Wehner[4] , and Hiroyuki Murakami[5]

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA, [2]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA, [3]Swiss Re, Zurich, Switzerland, [4]Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [5]Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

Abstract This manuscript discusses the challenges in detecting and attributing recently observed trends in the Atlantic tropical cyclone (TC) and the epistemic uncertainty we face in assessing future risk. We use synthetic storms downscaled from five CMIP5 models by the Columbia HAZard model (CHAZ), and directly simulated storms from high-resolution climate models. We examine three aspects of recent TC activity: the upward trend and multi-decadal oscillation of the annual frequency, the increase in storm wind intensity, and the decrease in forward speed. Some data sets suggest that these trends and oscillation are forced while others suggest that they can be explained by natural variability. Projections under warming climate scenarios also show a wide range of possibilities, especially for the annual frequencies, which increase or decrease depending on the choice of moisture variable used in the CHAZ model and on the choice of climate model. The uncertainties in the annual frequency lead to epistemic uncertainties in TC risk assessment. Here, we investigate the potential for reduction of these epistemic uncertainties through a statistical practice, namely likelihood analysis. We find that historical observations are more consistent with the simulations with increasing frequency than those with decreasing frequency, but we are not able to rule out the latter. We argue that the most rational way to treat epistemic uncertainty is to consider all outcomes contained in the results. In the context of risk assessment, since the results contain possible outcomes in which TC risk is increasing, this view implies that the risk is increasing.

Plain Language Summary We use a set of computer model simulations to study recent trends in Atlantic tropical cyclones. We looked at three aspects of these storms: the number of tropical cyclones each year, which has fluctuated up and down over time (but generally increased over the last several decades); the strength of their winds, which has been increasing; and the speed at which they move, which has been decreasing. These trends could be caused either by human-induced global warming or by natural variability; determining which cause is more important to overall risk requires us to understand how the number of tropical cyclones per year responds to warming. In our simulations, this number can either increase or decrease with warming, depending on which of two nearly identical versions of our model we use to simulate the storms. This uncertainty prevents us from reaching definitive conclusions about either present or future hurricane risk. Nonetheless, our analysis suggests that the risk of Atlantic tropical cyclones is more likely increasing than decreasing, and we argue that from a broader point of view, this is effectively equivalent to saying the risk is increasing.

1. Introduction

Rational measures to mitigate any risk must start from an assessment of that risk. Historical records can provide guidance, but in the case of atmospheric hazards such as hurricanes, we know that historical records are only a starting point for assessing current and future risk. This is both because the historical record is too short to fully sample the possibilities even in a stationary climate, and because the climate is changing (D. Chan et al., 2022; Emanuel, 2021; Schreck et al., 2014). Climate change makes the present different from the past, and requires us to consider whether the historical record alone—or catastrophe models that are built upon it—using purely statistical methods and assuming a stationary climate, are adequate, or need to be modified or supplemented to account for climate change.

Writing – original draft: Chia-Ying Lee
Writing – review & editing: Chia-Ying
Lee, Adam H. Sobel, Michael K. Tippett,
Suzana J. Camargo, Marc Wüest, Michael
Wehner, Hiroyuki Murakami

Accounting for climate change is likely to require a greater use of physics than is historically typical in catastrophe models (Emanuel, 2008; Toumi & Restell, 2014). While one might instead try to assess the risk by using standard statistical methods but training only on the most recent observations (as opposed to the entire record), in the hope that those most recent observations represent the present and near-future climate adequately, this is likely to be challenging. Since hurricanes are rare, the number in the record over a period recent enough for this purpose is too small for risk assessment—especially when we also consider that low-frequency natural variability is present (i.e., J. C. Chan, 2008; Klotzbach & Gray, 2008; Wang et al., 2015), so that averaging times must be longer than might otherwise be necessary. To make the best possible assessment of present hurricane risk, then, we need to use our knowledge of the physics that connects hurricanes and climate (Emanuel, 2008).

The focus of this study is Atlantic tropical cyclone (TC) risk in the present and near future. Future projections are useful for understanding how TCs may respond to climate changes of various sorts. Studies of historical observations, on the other hand, often look for trends; but on their own, such studies do not establish the causes of the trends, nor whether they will persist. Establishing whether a trend is present (detection) is generally viewed as a prerequisite to determining its cause (attribution) (Lloyd & Oreskes, 2018). Detection can, in principle, be done with observations alone; attribution requires a model of some sort, in order to construct a counterfactual where the cause of interest is not present (Hegerl & Zwiers, 2011; Knutson, 2017). If a historical trend (or an oscillatory signal) could be both detected and attributed to a specific cause, such as human influence, or alternatively some specific natural process, this would be of great scientific value, and would also allow us some insight into what to expect in the near future.

To develop such insight for Atlantic TCs, we will use recent observations and model simulations from historical (present), projected near future (up to 2040 or 2050), and pre-industrial periods. Simulations from the pre-industrial period incorporate no anthropogenic forcing and thus are used as a counterfactual. We use two types of model data. The first represents synthetic storms generated from a statistical-dynamical model, the Columbia (TC) HAZard model (Columbia HAZard model (CHAZ)), a model that encodes physical relationships between tropical cyclones and their ambient large-scale environment (Lee et al., 2018). The second represents the directly simulated hurricanes from high-resolution global models, in which the above-mentioned relationships are simulated intrinsically (Roberts et al., 2020a, 2020b; Wehner et al., 2014; Yoshida et al., 2017).

There are three objectives of this work. The first is to examine whether recently reported trends can be attributed to anthropogenic forcing. As summarized in Knutson et al. (2020a, 2020b), these trends are the recent variability of Atlantic annual TC frequency (Emanuel, 2007), an upward trend in the intensification rate (Bhatia et al., 2019) and lifetime maximum intensity (LMI) (Kossin et al., 2013), and a slowing-down in the storm motion (Kossin, 2018). In particular, the cause of the recent increasing trend (since 1970) in Atlantic TC activity has been a subject of debate. On the one hand, reduced aerosol forcing over the Atlantic since 1980s has been argued to be a dominant cause of the increasing TC activity in late 20th century (Mann & Emanuel, 2006; Rousseau-Rizzi & Emanuel, 2020; Sobel, Camargo, & Previdi, 2019). On the other hand, several measures of Atlantic TC activity, including the major hurricane (TCs with LMI ≥93 kt) frequency (Goldenberg et al., 2001), are highly correlated to the Atlantic Multi-decadal Oscillation (AMO) or Atlantic multidecadal variability, a low-frequency mode of variability identified by the average sea surface temperature (SST) anomalies in the North Atlantic basin, typically over 0–80°N (Ting et al., 2011). The recent AMO cycle, including both the upward trend from 1970 to 2005 and the downward trend from 2006 to 2018 have been associated by some authors with natural variability (e.g., Yan et al., 2017, and others). However, studies using CMIP5 historical runs simulated an ensemble-mean AMO that is significantly correlated with the observed AMO, suggesting that the recent historical variability could be a consequence of radiative forcing (Bellomo et al., 2018; Clement et al., 2015). The future projections of TC frequency are subject to a similar degree of debate. Many studies have suggested that the future should see a decline in the numbers of Atlantic TCs with warming (e.g., Knutson et al., 2010, and others), with a few exceptions (Bhatia et al., 2018; Emanuel, 2013; Vecchi et al., 2019).

The second objective is to compare historical simulations with observations to understand which modeling data set is more consistent with the observations (Brunner et al., 2020). Such analysis can provide guidance whether to favor one model over another, which is especially useful for reducing uncertainty when the projections cover a wide range even with an opposite sign, such as the projections of the divergent scenarios in global TC genesis (i.e., Lee et al., 2020; Sobel et al., 2021). Lastly, we will assess hurricane risk over a set of selected gates, which can be a sections of coastline or area of interest (represented using line gates) in the present and future climates.

Strictly speaking, risk includes severity of the hazard, exposure, and vulnerability of the properties of interest. Only the hazard component is examined here.

2. Data, Experimental Design and Method

2.1. Tropical Cyclone Data Sets

2.1.1. Observations

For reference, we use 6-hourly storm positions (longitude and latitude) and maximum wind speeds (in knots) from best-track data generated by the National Hurricane Center and obtained via the International Best Track Archive for Climate Stewardship v04r00 (IBTrACS Knapp et al., 2010) from 1951 to 2020. IBTrACS data have some limitations. First of all, due to the limited availability of satellite imagery and the Dvorak Technique for systematic image analysis, the IBTrACS data set for the 1950s and 1960s may have missed some storms (Vecchi & Knutson, 2008). Second, the intensity of extreme hurricanes may be underestimated from the 1940s to the 1960s (Hagen & Landsea, 2012). Furthermore, advancements in observational capabilities (Villarini et al., 2011) have led to an increase in the detection of short-duration and typically weaker tropical storms (Landsea et al., 2010). To address these limitations, we only consider storms, including subtropical cyclones, that reached tropical storm strength, with a lifetime maximum intensity (LMI) of at least 34 knots, and that stayed at 34 knots and above for a duration of at least 2 days (excluding weak and short-duration storms). When conducting trend analysis (in Section 2.4 and Section 3), we discuss uncertainties associated with observations using more recent data (1960–2020 and 1970–2020). It should be noted that the modeling data discussed below also incorporate the intensity and duration threshold to exclude short-duration storms. Throughout this study, the term "tropical cyclones" refers to storms with a LMI of at least 34 knots, while "hurricanes" are storms with a LMI of at least 64 knots. Storm forward motion is derived from position data.

2.1.2. Synthetic Events From the CHAZ Model

The first set of model TCs used here consists of synthetic storm tracks from the Columbia (TC) Hazard (CHAZ) model (Lee et al., 2018). CHAZ is a statistical-dynamical downscaling model that generates synthetic storms whose properties depend on environmental conditions. The environmental conditions can come from an observation-based reanalysis or a global climate model. The downscaling is one-way; there is no feedback from downscaled TC activity to the global models. Three components in CHAZ describe storm formation and subsequent evolution until shortly after landfall (or dissipation): the cyclone genesis index (TCGI; Tippett et al., 2011), the beta-advection track model (Emanuel, 2008), and an auto-regressive intensity model (Lee et al., 2015, 2016). Details about CHAZ are reported in Lee et al. (2018). The environmental variables required by the model are Potential Intensity (Bister & Emanuel, 1997), deep-layer (850–250 hPa) vertical wind shear, and one or more moisture variables: column integral relative humidity (CRH) and/or column integral saturation deficit (SD), the absolute vorticity at 850 hPa, and the steering flow. The choice of moisture variables will prove particularly important in what follows. Both variables are calculated following Bretherton et al. (2004). The simulated TC activity in CHAZ, at global and basin scales, in both current and projected future climates have been discussed in detail in Lee et al. (2018, 2020), respectively. The CHAZ model has been used for case studies in Texas (Hassanzadeh et al., 2020), New York (Lee et al., 2022), Mumbai, India (Sobel, Lee, et al., 2019) and the Philippines (Baldwin et al., 2023). Meiler et al. (2022) found that losses estimated from CHAZ are comparable to those estimated using comparable TC hazard models from Emanuel (2013) and Bloemendaal et al. (2020).

In this study, we use CHAZ to downscale five CMIP5 models (Taylor et al., 2012) over the Atlantic basin. They are the National Center for Atmospheric Research Community Climate System Model 4 (Gent et al., 2011), the Geophysical Fluid Dynamics Laboratory Climate Model version 3 (Donner et al., 2011), the United Kingdom Meteorological Office Hadley Center Global Environment Model version 2 Earth System (HadGEM2-ES) (Jones et al., 2011), the Max Planck Institute Earth System Model Medium Resolution (MPI-ESM-MR) (Zanchettin et al., 2012), and the Model for Interdisciplinary Research Climate Version 5 (Watanabe et al., 2010) from the University of Tokyo Center for Climate System Research, National Institute for Environmental Studies, Japan, Japan Agency for Marine-Earth Science.

CHAZ's projections of annual TC frequency, both in the Atlantic and globally, are sensitive to whether CRH and SD are used in TCGI. Using TCGI with CRH leads to a projected increase in global (and Atlantic) TC

frequency, while SD leads to a projected decrease (Lee et al., 2020). CRH and SD both measure the degree of the saturation of the atmosphere, with SD being the difference between the column integrated water vapor and the same quantity at saturation, and CRH being their ratio. As saturated water vapor increases with temperature in a warming climate, CRH remains close to constant and SD decreases (defined as negative, Camargo et al., 2014; Lee et al., 2020). In the current climate, however, the behavior of these two variables are qualitatively similar, and the two TCGI formulations yield similar results for the historical period, meaning that the historical evidence is inadequate to determine which of the two is more correct. Arguably, SD better reflects the increase in the thermo-dynamic inhibition of TC formation in a warming climate (Emanuel, 1989, 2022), but the gaps in our understanding of the relationship between climate and TC frequency are so substantial that we do not view this argument as dispositive (Sobel et al., 2021). The diverging annual frequency projections from CHAZ thus, in our view, reflects the broader state of the science, in that we have low confidence regarding whether one should expect more or fewer hurricanes as climate warms (i.e., Camargo et al., 2020; Vecchi et al., 2019; Sugi et al., 2020). One reason for the low confidence in TC frequency projection is the lack of theoretical understanding of TC genesis, and we refer the readers to a review article by Sobel et al. (2021) for a detailed discussion.

Since total TC hazard and risk depend inextricably on TC frequency and we lack a strong basis for choosing between SD and CRH, the sensitivity to the humidity variable in our results causes a deep uncertainty in the projected risk. This uncertainty will remain in the present study, in that we performed separate sets of simulations with either CRH or SD as the humidity variable in the genesis module, referred to as $CHAZ_{CRH}$ and $CHAZ_{SD}$.

2.1.3. Directly Simulated Hurricanes From General Circulation Models

In addition to the CHAZ downscaling simulations described above, we use storms tracked in a set of relatively high-resolution, that is, TC-permitting, global climate models. The first one is the 60-km MRI-AGCM3.2H large-ensemble simulation from Mizuta et al. (2017) (MRI-LENS). Tropical cyclones in that model were tracked following Murakami et al. (2012). Thresholds of sea level pressure, 850 hPa relative vorticity, 850 hPa and surface wind speed, warm core temperature, and duration period for the TC tracking were chosen in accordance with the model characteristics and resolution (see Section 2.4 of Murakami et al. (2012) for detail). Yoshida et al. (2017) demonstrated that the MRI-LENS model successfully simulates an annual global TC frequency of 83.2 cyclones per year and provides a reasonable global representation. However, the model underestimates the frequency of TCs in the North Atlantic region. The second one is the 25-km High-Resolution Community Atmospheric Model version 5 simulations, CAM5 (Wehner et al., 2014, 2015). TCs in CAM5 are tracked with Geophysical Fluid Dynamics Laboratory tracking algorithm using threshold values for vorticity, warm core temperature anomaly, and planetary boundary layer depth as in Knutson et al. (2007). Details of the tracking can be found in Appendix B in Wehner et al. (2014). Similar to the MRI-LENS model, the 25-km High-Resolution CAM5 produced an average of approximately 83 TC per year globally. This model is capable of simulating category 5 tropical cyclones on the Saffir-Simpson scale, and its representation of Atlantic tropical cyclones and seasonal cycles aligned well with observations. Next, we use storms tracked in the Coupled Model Intercomparison Project Phase 6 (Eyring et al., 2016) High Resolution Model Intercomparison Project (HighResMIP) (Haarsma et al., 2016). Following Roberts et al. (2020a, 2020b) and Roberts et al. (2020a, 2020b), we use storms from CMCC-CM2 (Cherchi et al., 2019), CNRM-CM6 (Voldoire et al., 2019), EC-Earth3P-HR (Haarsma et al., 2020), HadGEM3-GC3.1 (Roberts et al., 2019), and MPI-ESM1.2 (Gutjahr et al., 2019). There are two HighResMIP configurations, one is forced with prescribed SST while the other is fully coupled. We only use the simulations from the fully coupled configuration which allows natural variability to occur freely during the historical period. To understand the sensitivity of model performance to the TC trackers, HighResMIP storms are tracked by TRACK (Hodges et al., 2017) and TempestExtremes (Ullrich & Zarzycki, 2017; Ullrich et al., 2021; Zarzycki & Ullrich, 2017), and both event sets are used here (see Appendix B in Roberts et al. (2020a, 2020b) for details). For convenience, we label modeled TCs from HighResMIP tracked with TempestExtremes as Hi-TempExt and those tracked with TRACK as Hi-TRACK.

2.2. Experimental Design

Except in MRI-LENS and CAM5, we use model TCs from the historical, near-term future, and pre-industrial control (piC, no anthropogenic forcing) scenario simulations. Note that the time range covered in each period varies by model. For the historical period, they are 1951–2005 for $CHAZ_{CRH}$ and $CHAZ_{SD}$, 1950–2010 for MRI-LENS, 1996–2016 for CAM5, and 1951–2014 for the two HighResMIP data sets. In the future period, $CHAZ_{CRH}$ and $CHAZ_{SD}$ contain storms from 2006 to 2040 under Representative Concentration Pathway 8.5 (rcp8.5) while HighResMIP storms are

Table 1
Data Characteristics

Data	Global climate models	Resolution	Ens	Period	Annual frequency	R[%]
IBTrACs	N/A	N/A	N/A	1951–2020	9.26	46.3
CHAZ$_{CRH/SD}$	HadGEM2_ES	N/A	100		8.8/15.9	26.5/54.7
	CCSM4				11/16.1	4.48/40
	GFDL_CM3			1951–2005; 2006–2040; piC	16.5/19.1	30.6/63.8
	MPI_ESM_MR				29.3/39.4	7.4/76.5
	MIROC5				11.9/18.3	14.23/46.6
MRI-LENS	MRI–AGCM3.2H	60 km	100	1950–2010; piC	2.3	1.71
CAM5	CAM5	28 km	5	1996–2005; piC	10.9	9.54
Hi-TRACK	CMCC-CM2-VHR4 (r1i1p1f1)	25 km	7		5.0	16.4
	CNRM-CM6-1-HR(r1i1p1f2)	50 km			21.0	52.4
	EC-Earth3P-HR (r1i1p2f1)	50 km			6.8	11.0
	EC-Earth3P-HR (r2i1p2f1)	50 km		1951–2014; 2015–2040; piC	6.5	22.3
	HadGEM3-GC31-HH (r1i1p1f1)	50 km			21.5	15.2
	HadGEM3-GC31-HM (r1i1p1f1)	50 km			19.7	13.2
	MPI-ESM1-2-XR (r1i1p1f1)	50 km			4.5	15.7
Hi-TempExt	CNRM-CM6-1-HR(r1i1p1f2)	50 km	6		13.4	14.8
	EC-Earth3P-HR (r1i1p2f1)	50 km			2	5.82
	EC-Earth3P-HR (r2i1p2f1)	50 km		1951–2014; 2015–2040; piC	2	5.1
	HadGEM3-GC31-HH (r1i1p1f1)	50 km			13.3	12.7
	HadGEM3-GC31-HM (r1i1p1f1)	50 km			12.4	0.56
	MPI-ESM1-2-XR (r1i1p1f1)	50 km			0.63	4.8

from 2015 to 2050 under Shared Socioeconomic Pathways5-85 (ssp585). Both are high-emission scenarios with an additional radiative forcing of 8.5 W m^{-2} by the year 2100 (Riahi et al., 2017) which considers a fossil-fueled development. High-emission scenarios are employed to enhance the detection of the potential climate change signal, although the scenario makes little difference for the first few decades (Intergovernmental Panel on Climate Change (IPCC), 2014; Hawkins & Sutton, 2009; Blanusa et al., 2022). Warming climate simulations for MRI-LENS and CAM5 are under a 4°C (Yoshida et al., 2017) and 1.5°C warming (Wehner et al., 2018) scenarios and thus are not used here. In piC, the labeling of year is arbitrarily in all data sets as all years are equivalent. The MRI-LENS and CAM5 piC simulations are exceptions. In MRI-LENS and CAM5, the observed SST information is prescribed in both historical and piC simulations as a lower boundary condition, but the long-term linear trend is removed in the piC simulations. In other words, MRI-LENS and CAM5 piC simulations still contain observed variations in SST. The piC simulations in MRI-LENS, called "no-warming" in Mizuta et al. (2017) and those in CAM5, following "Nat-Hist" in Stone et al. (2019), are designed with an underlying assumption that only the linear trend is anthropogenically forced, not the variability. This assumption, as we will discuss in the next Section, is debatable.

In each period, the CHAZ model was used to generate 20 track ensemble members per CMIP5 model and each track has 40 intensity ensembles (100 CMIP5 track ensemble members and 4,000 considering intensity ensemble), as is possible because the CHAZ intensity module has a stochastic component. Hi-TRACK has seven members (five global climate models and two of them have two ensemble members) and Hi-TempExt has six (four global climate models and two of them have two ensemble members). MRI-LENS has 100 ensemble members while CAM5 has five. The data properties are listed in Table 1.

2.3. Frequency Adjustment

There are biases in the statistics of all model TCs, because of biases in the models that generate them, including the CHAZ model itself as well as the CMIP5 models from which CHAZ obtains its environmental conditions,

and the high-resolution global climate models used here. In particular, all models have biases in TC frequency (Table 1), and directly simulated hurricanes from high-resolution global climate models have low biases in intensity, in general, as the grid spacings of these models are too coarse to capture the full range of observed hurricane strengths (e.g., Moon et al., 2022; Yoshida et al., 2017, and others). Here we address only the frequency biases. Specifically, we derive an adjustment by comparing the basin-wide annual TC frequencies of models' historical simulations to that of the observations from the same period. The same adjustment is then applied to both historical and future simulations. Similarly, we compare the annual frequency of the piC simulations to the observations to adjust piC's annual frequency. In Lee et al. (2018, 2020), the basin-wide frequency adjustment is a multiplicative factor to ensure that the mean annual frequency over a basin in CHAZ is consistent to that in observations. However, some ensemble members of high-resolution global climate models used here, such MRI-LENS, generate zero TCs in some years. A multiplicative factor would result in larger variability but still have zeros in these years, which is unrealistic. Thus, here the basin-wide frequency is adjusted as:

$$f_{\text{adj}} = \sigma_{\text{obs}} \times \frac{f_{\text{ori}} - \mu_{\text{model|ref}}}{\sigma_{\text{model|ref}}} + \mu_{\text{obs}}, \tag{1}$$

where f indicates annual frequency (each year) with the subscript indicating after ($_{adj}$) or before ($_{ori}$) frequency adjustment. μ and σ are the mean and standard deviation of the frequency and the subscript indicates whether it is from simulations ($_{model}$) or observations ($_{obs}$). As we want to retain the climate change signal, reference μ and σ ($\mu_{model|ref}$ and $\sigma_{model|ref}$) for adjusting frequencies in both historical and future simulations in each data set are from its respective historical simulation. μ_{obs} and σ_{obs} are calculated using observations from each data set's respective historical period. To adjust the annual frequencies of the piC simulations, $\mu_{model|ref}$ and $\sigma_{model|ref}$ are from piC. Biases in annual TC frequency of the piC simulations are different to those in the historical simulations. As we will discuss later, a basin-wide frequency adjustment may not correct regional biases, because model biases can have spatial dependence. When desired, that is, at a local gate for regional risk assessment (in Section 5), we apply a multiplicative factor so that the annual frequency of simulated storms with intensity greater than 40 kt match that of observations, which is the same as the bias-correction approach used in Lee et al. (2022).

Although the model TC data sets utilized here exhibit biases, each data set has undergone comprehensive evaluation as documented in Section 2.2 and Section 2.3 and are suggested to be adequate for investigating TCs and climate change. The application of frequency adjustment in this study represents our effort to optimally utilize these data sets. An underlying assumption of our approach to bias correction, in common with many climate change studies, is that the bias of any given model remains the same in projected future climate periods as it is in the present, so that the influence of the projected climate change can still be captured when comparing simulations between future and historical periods. This assumption is analogous to that used to remove climatological biases in surface temperature and other quantities from the climate models themselves in global warming projections, for example, those by the Intergovernmental Panel on Climate Change (Solomon et al., 2007). While this assumption of constant biases can be questioned, it is a simple assumption, and there is no empirical basis on which to base any more complex assumption one. Still, we will discuss the impacts of frequency adjustments on our findings.

2.4. Trend Analysis

To calculate trends of TC activity, we fit second-order Legendre polynomials:

$$\hat{y} = a_0 + a_1 x + \frac{a_2}{2}(3x^2 - 1), \quad x \in [-1, 1] \tag{2}$$

to the time series of the variables of interest from observations and model simulations. In Equation 2, x is years scaled to interval of $[-1, 1]$, $\hat{y}$ represents the fitted variables, the coefficient a_1 shows linear trends and a_2 shows quadratic trends. Considering quadratic trends allows the possibility that the observed multi-decadal variability is in fact forced (Bellomo et al., 2018; Clement et al., 2015). Here, we ask whether or not the observed trends lie within the ensemble spread from simulations. For example, if the observed trend is outside of the range of piC simulations but is within those from historical simulations, then the observed change (e.g., upward trend or increasing curvature) is unlikely to have occurred without anthropogenic forcing. When comparing the trends between observations and simulations, a_1 and a_2 are scaled back so that they have units of the variable's unit per year (yr^{-1}) and per year square (yr^{-2}), respectively.
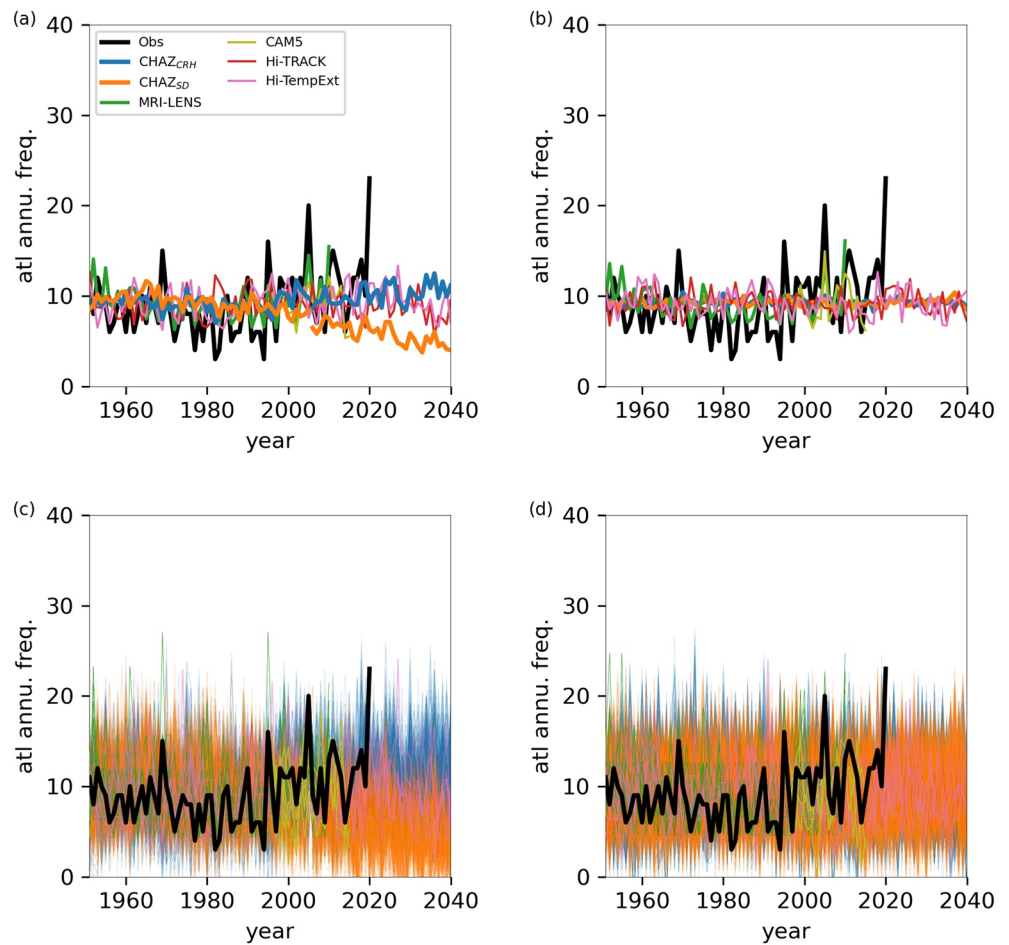
Figure 1. Annual frequency of Atlantic TCs which lifetime maximum intensity reach at least 34 kt intensity threshold from 1951 to 2020 from best-track data (black), CMIP5 downscaling simulations using $CHAZ_{CRH}$ (blue) and $CHAZ_{SD}$ (pink), 25-km high-resolution CAM5 simulations (purple), 60 km MRI-LENS simulations (green), and High Resolution Model Intercomparison Project (HighResMIP) simulations from Roberts et al. (2020a, 2020b) and Roberts et al. (2020a, 2020b). Storms from HighResMIP are tracked with TRACK (red) and TexmpExtreme (pink), respectively. In (a) and (c), simulations in their respective historical period are conducted with historical climate forcing while those in future period are with the rcp8.5 (for Columbia HAZard model) and ssp585 (for HighResMIP) warming scenarios. In (b) and (d), the simulations are under pre-industrial control climate (no anthropogenic forcing). (a) and (b) show the results from ensemble mean while (c) and (d) show the results from all ensemble members.

3. Trend and Multi-Decadal Variability

3.1. Atlantic TC Frequency

We first examine the Atlantic TC frequency trends in the historical (present) climate and from historical to the warming future (i.e., using simulations from both historical and future periods). Figures 1a and 1b show the ensemble means of the time series of Atlantic TC frequency, that is, the averaged total number of storms in the basin each year whose maximum sustained winds reach at least 34 kt from each data set. The small wiggles may be sampling variability. Figures 1c and 1d show the ensemble spread. By construction, the time-mean annual frequency for each data set over its respective historical period will be identical to observations after frequency adjustment (Equation 1). The original annual frequency of each data set is shown in Table 1. Before 2000, the different simulations are, by eye at least, indistinguishable in their overall envelopes, with none showing any particular trend, and the observations (black thick line) lying well within their spread (shown in Figure 1c). After 2000, the $CHAZ_{SD}$ (orange thick line) and $CHAZ_{CRH}$ (blue thick line) results begin to diverge, with $CHAZ_{SD}$ showing a decreasing TC frequency and $CHAZ_{CRH}$ showing an increasing TC frequency. It is possible that this is related to the fact that the rcp8.5 scenario starts after 2005. The two HighResMIP data sets show no considerable

trend in the historical period but a sharp dip after 2030. The ssp585 scenario in HighResMIP starts after 2015, though. Hi-TRACK's annual TC frequency climbs up by 2040. Roberts et al. (2020a, 2020b) reported that both Hi-TRACK and Hi-TempExt project a reduction of ensemble mean annual frequency (less than 10%) from 1950–1980 to 2020–2050, but the spread covers zero, indicating low confidence in the mean trend.

Figures 1b and 1d show analogous results for piC simulations. Note that the years in the *x*-axis do not correspond to the actual historical years for most models; these labels are placed so we can compare the simulated trends to the observed trend and those in Figures 1a and 1c. Two exceptions are MRI-LENS and CAM5 simulations; both are uncoupled atmospheric models and forced with observed SST with anthropogenic trend removed (See Section 2 for details). In Figure 1b, the CHAZ$_{CRH}$ and CHAZ$_{SD}$ results do not diverge. There is no dip in the Hi-TRACK or Hi-TempExt. Clearly, the separation between CHAZ$_{CRH}$ and CHAZ$_{SD}$ and the dip in the two High-ResMIP data sets in Figure 1a represent forced responses.

Next we conduct the trend analyses of the annual TC frequency in Figure 1 using second-order Legendre polynomial fits (Equation 2). As an example, Figure 2a shows the analysis using the CHAZ$_{CRH}$ simulations and the observations. The observed fit (dashed black line) has an upward trend of 0.061 storm year^{-1} and a positive curvature of 0.0036 storm year^{-2} (shown as the first black dot and the horizonal black line in Figures 2b and 2c). The existence of a linear trend means that there is an overall increasing trend in storm activity since 1951 while the quadratic terms captures the multi-decadal variability, with high activity in the 1950s–1960s, low in the 1970s–80s, and high after that, which recent research suggests may be a forced signal rather than natural variability (Bellomo et al., 2018; Clement et al., 2015). The 90% confidence intervals depicted in Figures 2b and 2c consistently lie above the zero line, indicating a reasonable level of statistical significance for both the upward trend and positive curvature observed. To account for the potential impact of missing data in the 1950s and 1960s discussed in Section 2.1, we applied second-order Legendre polynomial fits to the data starting instead in 1961 and 1971, represented by the second and third black dots in Figures 2b and 2c. These fits reveal stronger upward trends, measuring 0.09 storms year^{-1} and 0.15 storms year^{-1} for the periods of 1961–2020 and 1971–2020, respectively. The positive curvatures are determined as 0.004 storm year^{-2} and 0.003 storm year^{-2}. However, when excluding the high activity in the 1950s–1960s, the 90% confidence interval for positive curvature derived using 1971–2020 data includes the zero line, indicating no statistical significance at the 10% level. We also conducted Legendre polynomial fits for Atlantic hurricanes (with LMI greater than 64 kt) frequency using both the IBTrACS data set and the adjusted records from Vecchi et al. (2021) that account for missing storms. The derived trend and curvature from IBTrACS are 0.03 storms year^{-1} and 0.00175 storms year^{-2}, respectively, while they are 0.02 storms year^{-1} and 0.0013 storms year^{-2} using the data from Vecchi et al. (2021). This exercise suggests that the missing storms in 1950s and 1960s may introduce slight variations in the absolute values, but do not cause drastic changes to the overall results. Hence, for the subsequent discussions, we use only the IBTrACS data.

In Figure 2a, the polynomial fits of CHAZ$_{CRH}$ simulations from historical only (light blue dashed line) and from historical to future (dark blue dashed line) simulations both show small upward curves while the polynomial fit derived from the piC simulations (gray dashed line) is quite flat. The ranges of the fit parameters from all ensemble members in each data set are also shown in Figures 2b and 2c. The means of simulated trend from individual approaches are smaller (in magnitude) than the observed trend, and some have negative signs (Table 2). For examples, the mean linear trends in the historical period are 0.00061, −0.035, 0.0089, −0.047, 0.01, and 0.012 for CHAZ$_{CRH}$, CHAZ$_{SD}$, CAM5, MRI-LENS, Hi-TempExt and Hi-TRACK, respectively.

As discussed in Section 2.4, a question we are asking here is whether not the observed trends lie within the ensemble spread from simulations. Figure 2b show that the observed linear trend is above most of the piC simulations except those from CAM5. However, CAM5 has only 10 years of simulations, which is too short to be compared with 70 years of observations. The observed quadratic term is outside of the ensemble ranges of piC simulations from CHAZ$_{CRH}$, CHAZ$_{SD}$, and the two HighResMIP data sets, and is within the 25–75 percentile ensemble ranges of MRI-LENS simulations and in the top 25th percentile (75–100 percentile) in CAM5. The observed linear trend is in the top 25th percentile (75–100 percentile) of the historical simulations of both CHAZ$_{CRH}$ and CHAZ$_{SD}$, MRI-LENS and CAM5, and is marginally covered by the simulations of Hi-Track; the observed quadratic term is within the top 25th percentile range of the historical simulations of CHAZ$_{CRH}$, CHAZ$_{SD}$, and CAM5, and is withing 25–75 percentile in MRI-LENS. Only the linear trend derived from historical + future simulations of the CHAZ$_{CRH}$ includes the observed value. For the quadratic trend, the observed values are above model fitted values. (We do not use any warming simulations from CAM5 and MRI-LENS.)
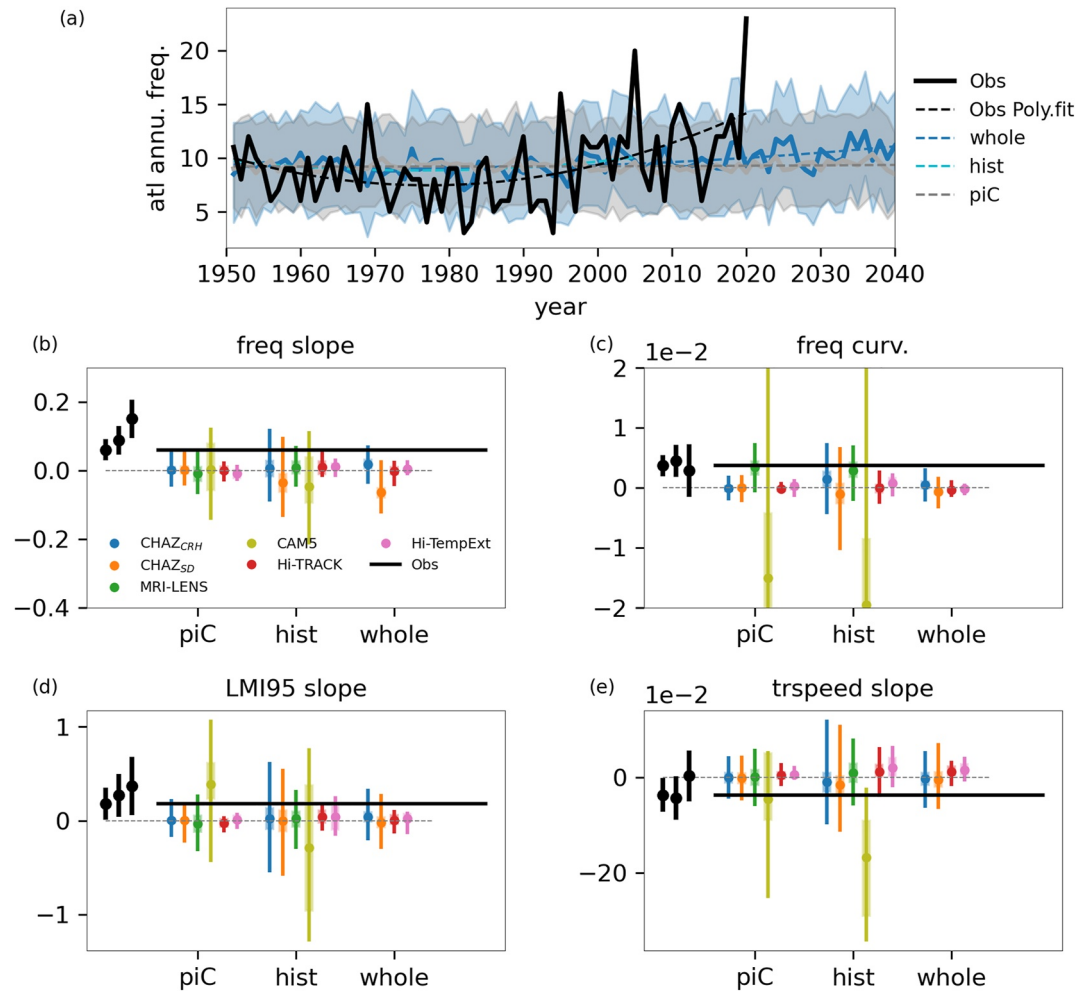
Figure 2. (a) Observed (black) and CHAZ$_{CRH}$ simulated mean annual tropical cyclone frequency. The CHAZ simulations are from present (1951–2005) to future climate (2006–2040) periods (blue), and from those using pre-industrial control climate forcing (gray). Dashed lines show the polynomial fit. "hist" shows the fit using synthetic storms from historical period only while "whole" are from the historical and future periods. (b) Linear terms of the polynomial fit derived using synthetic and directly simulated storms' annual frequency from all data sets. Data sets are indicated by color while the black line show the observed value. The three black dots (from left to right) and the vertical lines show the observed value and the 95% confidence interval level using data starting in 1951, 1961, and 1971. (c) Similar to (b) but for the quadratic terms. (d) and (e) are similar to (b) but for linear terms from the polynomial fit of LMI95 and storm translation speed. Units for (b), (c), (d) and (e) are, respectively, storm number year^{-1}, storm number year^{-2}, m s^{-1} yr^{-1}, and km hr^{-1} yr^{-1}.

Generally speaking, the polynomial fit analysis suggests that, first, CHAZ$_{CRH}$, CHAZ$_{SD}$ and MRI-LENS better capture the observed trend and multi-decadal variability than others. However, CAM5 has only 10 years of data with five ensemble members, and Hi-TRACK and Hi-TempExt have only, respectively, seven and six ensemble members. These three data sets may not be large enough to represent their own internal variability fully. Second, the observed linear trend is marginally covered by the spread of CHAZ$_{CRH}$, CHAZ$_{SD}$ and MRI-LENS' piC simulations but within the spread of these models' historical simulations, indicating that anthropogenic forcing is necessary to capture the upward trend in the past decades. On the other hand, we cannot rule out the possibility that the recent upward curvature is within the range of natural variability. The MRI-LENS' piC simulations are forced with the observed SST with long-term trend removed and thus have the upward curvature right on top of observed values in Figure 2c. Simulations from CHAZ$_{CRH}$ and CHAZ$_{SD}$ suggest that anthropogenic forcing helps to capture the upward curvature trend. Third, when considering the future period as well, the mean of CHAZ$_{CRH}$ shows an upward trend, the mean of CHAZ$_{SD}$ shows a downward trend, and the mean of the two HighResMIP simulations are close to zero. However, we have low confidence in the projected trend as their spread includes zero. Thus, we cannot say for sure that the positive linear and quadratic trends will continue into the future.

Table 2
Ensemble Mean of the Legendre Polynomial Fitted Parameters

Data	Period	Freq. Slope (storm number year^{-1})	Freq. Curv. (storm number year^{-2})	LMI95 slope (m s^{-1} yr^{-1})	Trspeed slope (km hr^{-1} yr^{-1})
IBTrACs	Historical	6.1E−2	3.6E−3	0.18	−3.6E−2
CHAZ$_{CRH/CRH}$	Historical	6.1E−3	1.4E−3	2.3E−2	−9.5E−3
	Whole	1.9E−2	5.0E−4	3.8E−2	−3E−3
	piControl	1.1E−3	−1.3E−4	1.1E−3	−1.0E−3
CHAZ$_{CRH/SD}$	Historical	−3.5E−2	−1.0E−3	−2.6E−3	−1.34E−2
	Whole	−6.3E−2	−6.4E−4	−1.9E−2	−3E−3
	piControl	1.6E−3	3.03E−5	1.6E−3	−2.0E−3
MRI-LENS	Historical	8.9E−3	2.8E−3	2.0E−2	7.37E−3
	Whole	N/A	N/A	N/A	N/A
	piControl	−9.0E−3	3.4E−3	−3.3E−3	6.0E−4
CAM5	Historical	−4.7E−2	−2.0E−2	−0.3	3.78E−2
	Whole	N/A	N/A	N/A	N/A
	piControl	3.7E−3	1.5E−2	3.9E−1	−4.6E−2
Hi-TRACK	Historical	1.0E−2	−4.2E−5	4.0E−2	9.0E−3
	Whole	−2.5E−3	−3.9E−4	4.6E−3	9.5E−3
	piControl	4.3E−4	−1.9E−4	−2.6E−2	4.3E−1
Hi-TempExt	Historical	1.2E−2	7.5E−4	3.9E−2	1.6E−2
	Whole	4.8E−3	2.2E−4	2.0E−2	1.21E−2
	piControl	−9.1E−3	2.5E−4	9.4E−3	5.1E−3

It should be noted that without the basin-wide frequency adjustment (not shown), the observed linear and quadratic trends lie outside of the spreads of MRI-LENS, Hi-TempExt and Hi-TRACK in all three periods. They are within the spreads of the CHAZ$_{CRH}$ and CHAZ$_{SD}$ simulations in the piC and historical periods. With additional data from 2006 to 2040, only CHAZ$_{CRH}$ shows such an upward trend continuing into the future.

3.2. Intensity and Storm Motion

Figure 2d shows the fit parameters of Atlantic TC LMI. Specifically, we look at the variability of the 95th percentile of LMI (LMI95), for which an upward trend has been found in observations (Kossin et al., 2013). Here we focus on the ensemble spread of the linear term only. The ensemble mean of the linear trend is shown in Table 2. There is an upward trend in the observations (0.18 m s^{-1} year^{-1}), meaning that the extreme tail of observed intensity has increased with time, consistent with previous studies (e.g., Knutson et al., 2020a, and others). Similar to Figure 2b, the upward trend of LMI95 is more evident when analyzing data starting in 1971. The positive linear trend is captured by the ensemble spreads of the two CHAZ data sets and those from MRI-LENS and CAM5 for both piC and historical periods. It is outside of the ensemble spread of all simulations from Hi-TRACK and Hi-TempExt. Thus, at least based on CHAZ$_{CRH}$, CHAZ$_{SD}$, MRI-LENS, and CAM5, we cannot rule out the possibility that the recent upward trend in the LMI95 is due to natural variability. When looking into the future, only the mean of CHAZ$_{CRH}$ spread is positive, and it is negative for CHAZ$_{SD}$. The mean of the spreads of Hi-TempExt and Hi-TRACK are close to zero. Similar to the results from TC frequency, the ensemble spread in Figure 2d include zero in the whole historical + future periods combined, indicating, again, low confidence in the directions of the projected changes.

Figure 2e shows the analysis for translation speed. Consistent with Kossin (2018), the observations show a clear downward trend (−0.036 km hr^{-1} year^{-1}). The 90% confidence interval of the observed trend includes zero, however. This observed trend is within ensemble spread in all periods, including piC, for all models, except the simulations from Hi-TempExt. However, the mean and the 25th–75th percentile ensemble spreads in these data sets move toward different directions from piC to historical to historical + future periods. The Hi-Track and

MRI-LENS HIST simulations show upward trends in the storm motion and these upward trends continues in to the future. The differences in mean and 25th–75th percentile ensemble spreads from CHAZ$_{CRH}$ and CHAZ$_{SD}$ from these three periods are small. The piC and HIST simulations from CAM5 suggest that anthropogenic forcing may lead to a strong downward trend in storm translation speed but again CAM5 simulations covers shorter period than are those of the other models. It seems unjustified, based on this set of models, to attribute the observed slowing down to anthropogenic forcing. It also noteworthy that at a regional scale, CHAZ projected an upward trend in storm translation speed for TCs affecting Texas (Hassanzadeh et al., 2020) and a downward trend for storms impacting New York (Lee et al., 2022). Spatially inhomogeneous changes may dilute the basin-wide signal.

4. Likelihood Comparison

Figure 2 shows that the simulated trends in historical and historical + future TC statistics vary from one data set to another. This is especially true for the different TC frequency projections between CHAZ$_{CRH}$ and CHAZ$_{SD}$, but a qualitatively similar result, including both increasing and decreasing trends, holds for the rest of our ensemble of opportunity. It is natural to ask whether we can develop some criteria for determining which is correct. In climate science, the multi-model ensemble mean is a common approach to obtain the consensus from multiple global climate models. However, such an approach is only adequate when the ensemble spread represents variations that can be considered random, as might be the case with typical aleatoric (stochastic) uncertainties. The divergent scenarios in the frequency projections are a consequence of the epistemic uncertainty (Epistemic uncertainty refers to the uncertainty stemming from limitations in scientific knowledge, models, and data used to project future climate conditions (Beven et al., 2018; Shepherd, 2019)) due to the lack of a satisfactory scientific understanding of TC frequency (Emanuel, 2022; Sobel et al., 2021) and thus the multi-model mean may not be meaningful in this case. We can, however, use likelihood analysis, in which the probabilities that the observations occurs within the model simulated distribution were computed. Thus, we can determine which simulation the observation is more consistent with. This is similar to the Likelihood Skill Score used for evaluating weather and climate predictions (Barnston et al., 2010).

Specifically, we first assume that annual hurricane frequency is drawn from a Poisson distribution whose mean (λ_t) has a trend in time ($\lambda_t = at + b$). We then obtain a and b of each data set by fitting the model annual TC frequency to a Poisson regression. We do so for all simulations (up to 2005 for CAM5 and 2010 for MRI-LENS). Note that with a and b, we can derive λ_t even for years beyond the data coverage period, that is, we can estimate f_{2020} with a and b derived from CAM5 data. The yearly likelihoods (L_t) of the observed frequencies are assigned based on the Poisson distribution with a given λ_t. For example, the likelihood CHAZ$_{CRH}$ simulations will generate 20TCs as observed in 2005 is 0.02%, which is based on a Poisson distribution with $\lambda_{2005} = 16.02$. The same calculation is applied to piC simulations, and the derived likelihood is denoted $L_{piC,t}$. For a given year, we then compare the log likelihood ratios L_t and $L_{piC,t}$ (i.e., $\log(L_t/L_{piC,t}) = \log(L_t) - \log(L_{piC,t})$). If this ratio is positive, the observations are more consistent with the simulations with anthropogenic forcing than with the piC simulations. If this ratio is negative, the observations are more consistent with the simulations with the piC simulations than with anthropogenic forcing.

We start by comparing the likelihoods of simulations with anthropogenic forcing to those with piC simulations (i.e., $\log(L_t/L_{piC,t})$ in Figure 3. The ratios of the likelihoods jointly up to 2020 (numbers on the upper-left in all panels) suggest that the observations are more consistent with the simulations with anthropogenic forcing than those without in CHAZ$_{CRH}$, MRI-LENS, and Hi-TempExt. The annual likelihood ratios from these three data sets further show higher annual likelihood (L_t) for the observed annual frequency values during the period of high TC activity in 1950–1970 and after 2000 while higher $L_{piC,t}$ is found during 1970–2000. This is because there are upward trends in the simulated annual frequency in these three data sets when compared to those in piC (Figure 2a). As λ_t increases with time, the distributions from these three data sets shift right with time and thus give greater likelihood to the high observed annual frequency when compared to those derived from piC simulations in which λ_t is close to constant in time. In contrast, CHAZ$_{SD}$ has a downward trend and its λ_t shifts left in time and leads to lower likelihood when observed values are high. Consequently, we see a higher $L_{piC,t}$ during high TC activity periods and higher L_t during the inactive TC seasons in CHAZ$_{SD}$. The frequency slopes obtained from piC and HIST in the Hi-TRACK data are similar and thus their log likelihood ratio is close to zero. When comparing the log likelihood ratio between every two data sets during 1951–2020 with anthropogenic forcing,
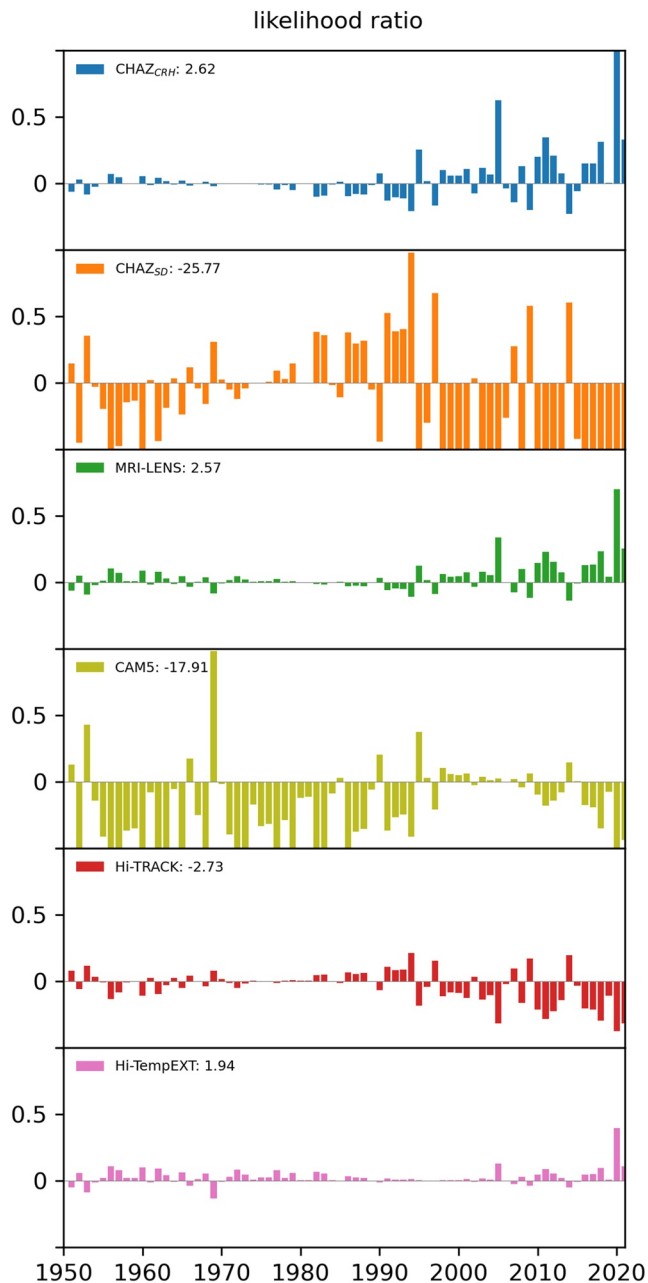
likelihood ratio

Figure 3. Annual log-likelihood ratio in which λ_t is derived from historical (and future for the Columbia HAZard model and High Resolution Model Intercomparison Project runs) simulations to that is estimated based on pre-industrial control simulations.

observations are more consistent to simulations in $CHAZ_{CRH}$ than other five data sets, followed by MRI-LENS, Hi-TempExt, Hi-TRACK, and CAM5.

The basin-wide frequency adjustment (Equation 1) that we performed to correct model biases is expected to affect the results of the likelihood analysis, because the frequency adjustment both shifts the mean of the model's TC annual frequency distributions and changes their shapes. The annual frequency distributions from historical and piC simulations are more distinct in the data sets without frequency adjustment, which indeed leads to larger log likelihood ratios (not shown). Without the frequency adjustment, the observed TC annual frequencies are more consistent with the historical simulations in $CHAZ_{CRH}$, $CHAZ_{SD}$, and MRI-LENS than in their respective piC simulations due to the large bias in these piC simulations.

While the Poisson distribution is commonly assumed for modeling TC frequency, TC data in practice may deviate from a Poisson distribution, which has a constant rate. For instance, in extreme years when environmental conditions, typically unfavorable for TC genesis, temporarily become favorable, the resulting peaks may not be well captured by the overall Poisson distribution. Also, pooling Poisson-distributed data with different rates results in data that are not Poisson distributed. In the case of $CHAZ_{CRH}$ and $CHAZ_{SD}$, for example, the genesis probability follows a Poisson distribution at each 2×2 grid, as described in Tippett et al. (2011). Each grid has its own mean rate, which depends on the local environment, meaning that different years, grid points, and even different CMIP5 models from which CHAZ downscaled yield different rates and tend to result in over dispersion (variance is greater than mean). And additional factor is that we apply a basin-wide frequency adjustment (Equation 1) that can also deviate the data deviate from a Poisson distribution. To mitigate impacts of over dispersion due to different models, we fit the Poisson regression and derived the likelihood for individual models. The likelihood reported in Figure 3 are the averaged values for $CHAZ_{CRH}$, $CHAZ_{SD}$, Hi-Track, and Hi-TempExt. The explained variance of historical + future TCs in each data sets is shown in Table 1. Using a negative binomial distribution, an alternative distribution that does not constrain the mean and variance to be equal, did not substantially change the results (not shown).

5. Climate Change and Regional Hurricane Risk at Three Line Gates

Now we compute regional hurricane risk, from hazard perspective only, represented by the return periods of storms of given wind intensities passing through pre-defined gates, derived from historical and future simulations. We use simulations from $CHAZ_{CRH}$, $CHAZ_{SD}$, Hi-TRACK, and Hi-TempExt. The three line gates used here (black lines in Figures 4a–4c) are: the main development region (MDR) gate, which can be thought of as delineating Atlantic TC hazard in a general sense, that is, how many storms form, and at what intensities and then move from the MDR toward the US and Caribbean Islands; the GoM gate, which records TC activity for those storms that enter the Gulf of Mexico; and the NE gate which is parallel to a portion of the Northeastern US coast. As discussed earlier (Section 2.3), to obtain more realistic return period curves for regional hurricane risk assessment, we use a more localized frequency adjustment. As an example, Figures 4d–4f show historical simulations from $CHAZ_{CRH}$ with basin-wide and regional frequency adjustments (Equation 1). While the basin-wide frequency adjustment (dashed lines) yields a TC frequency close to observations at the GoM gate, $CHAZ_{CRH}$ still overestimates storm activity at the MDR gate and underestimates storm activity at the NE gate. The regional frequency adjustments shift the simulated return period curves (solid line, local adjustment)
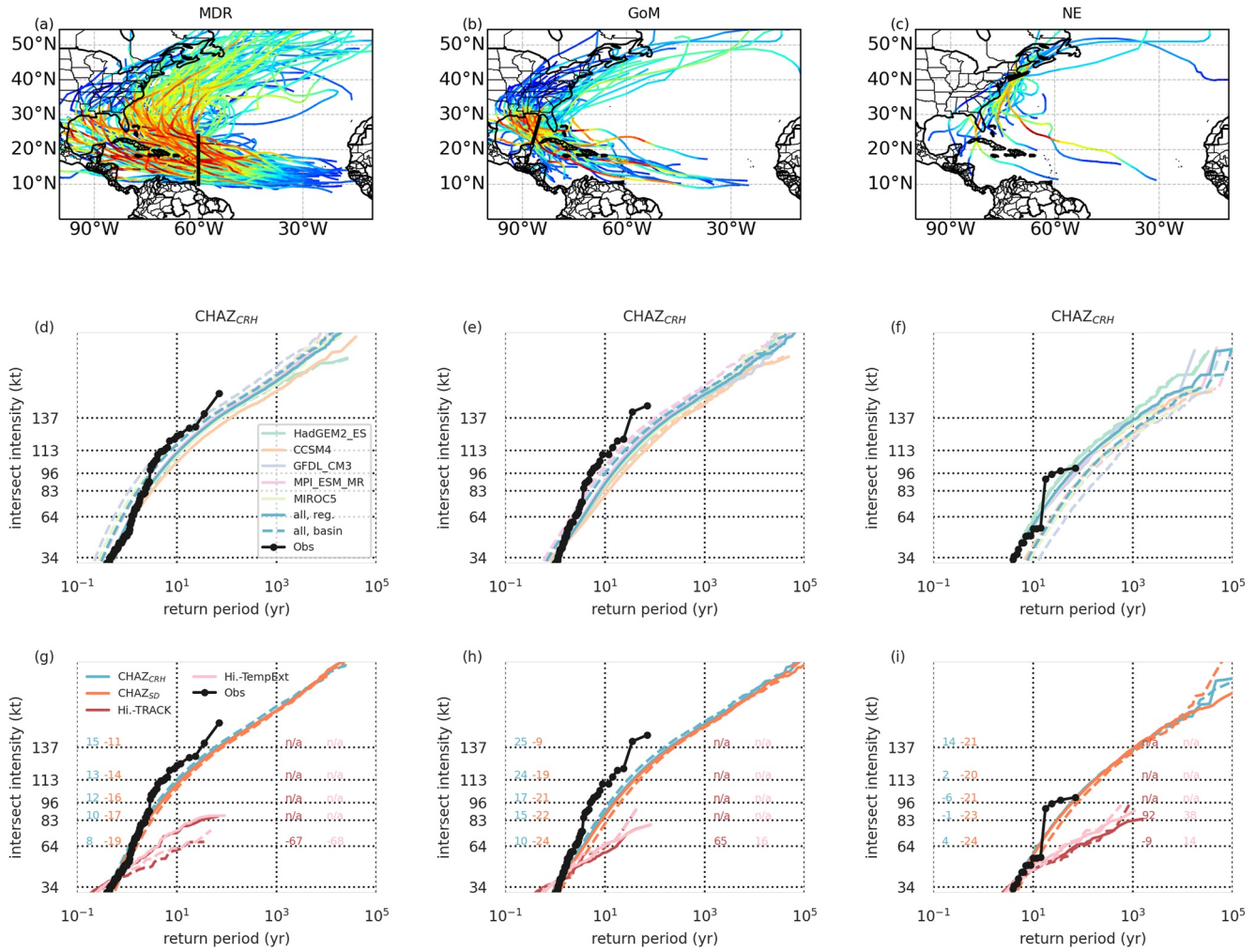
Figure 4. (a–c) Observed storm tracks from 1951 to 2020 at three line gates of interest. (d–f) Return period curves from 1951 to 2020 from best-track data (black lines), and CHAZ_CRH historical simulations with basin-wide (dashed lines) and local (solid lines) corrections applied at the three gates. Global climate model forcings are indicated by colors and blue lines show the derived return period curves using all data. (g–i) Similar to (d–f) but for the four data sets. The solid lines show the return period curves using all historical simulations while dashed lines use all future simulations. Numbers at each Saffir-Simpson intensity threshold are the percentage changes of the frequency of the storms exceeding the threshold. Data sets are indicated by colors. Black curves show the empirical return curve using observations from 1951 to 2020.

by matching the return periods at 40 kt to the values derived from observations (see Section 2.3 for details) at the gates. In terms of the shape of the return period curve, as well as the return periods at high intensities, CHAZ_CRH performs better at the MDR gate than at the GoM gate. It is difficult to directly compare the modeled curves to the observations at the NE gate, due to the significant underestimation of overall TC frequency at the latter. However, even there, the shapes of the observed and modeled return period curves are similar.

To show the changes in return periods between historical and future periods, Figures 4g–4i show the return period curves derived from the four data sets that have rcp8.5/ssp585 warming scenarios available. We use model storms from all ensemble members. Low-intensity biases in the Hi-TRACK and Hi-TempExt lead to an underestimate of the TC risk. HighResMIP models barely simulate storms with major hurricane wind strength (Moon et al., 2022; Roberts et al., 2020a, 2020b). The return period curves of CHAZ_CRH and CHAZ_SD HIST simulations are close to each other. The differences between simulations from historical period and those from historical and future periods, that is, the differences between the dashed and solid lines, are small for the two CHAZ data sets in Figures 4g–4i. Likewise the historical and future period curves of GoM and NE gates for Hi-TRACK and Hi-TempExt nearly indistinguishable. At the MDR gate, both Hi-TRACK and Hi-TempExt suggest increases in the TC risk.

To make these differences more evident, we list the percentage changes in annual TC frequency exceeding each Saffir-Simpson category on both sides of each panel in Figures 4g–4i. As expected, there is an overall increase in the storm frequency at all thresholds from historical to future periods for CHAZ$_{CRH}$ while there is an overall decrease for CHAZ$_{SD}$, consistent with the results in Figures 1 and 2a. The percentage changes are larger at higher intensity thresholds in the CHAZ$_{CRH}$ but they are of similar or smaller magnitude throughout the Saffir-Simpson categories in the CHAZ$_{SD}$. This is probably due to the increase in storm intensity as climate warms in CHAZ$_{CRH}$ and CHAZ$_{SD}$.

The changes in the frequency of exceedance at the three gates from Hi-TRACK and Hi-TempExt are not the same sign. Hi-TRACK shows a 67% decrease of Category 1+ ($\leq$64 kt) at the MDR gate but a 65% increase at GoM gate. At the NE gate, Hi-TRACK shows an 14% and 38% increase in the frequency of Category 1+ and 2+ hurricanes, respectively. Hi-TempExt shows a 68% decrease and 16% increase of Category 1+ hurricanes at the MDR and GoM gates, respectively. At the NE gate, it shows a 9% decrease and 92% increase in the frequency of Category 1+ and 2+ hurricanes. Storms from these two HighResMIP runs are undersampled and have low intensity biases (See Figure 7 in Roberts et al. (2019)). The directly simulated storms are not suitable for risk assessment and these numbers should be used with caution.

6. Discussion

The results of this study lead us to a view of Atlantic TC risk which requires us to confront epistemic uncertainty. We have multiple sets of simulations which give different views of the risk, in particular more so as we look further into the future. TC frequency increases in CHAZ$_{CRH}$ simulations and decreases in CHAZ$_{SD}$, a difference that hangs on a subtle modeling choice (SD vs. relative humidity as a predictor of genesis). Changes in the high-resolution global climate model simulations are smaller, but again depends on which models are considered.

The differences among these simulations are manifest not just in the future, but also to some degree in the present, and the observations do not allow us to determine with complete certainty which is correct. At present, no rigorous justification can be given regarding which choice to make. Thus, all these outcomes—increasing, decreasing, and no change in TC frequency in response to radiatively forced warming—have to be treated as possible. One may favor a data set over the others following the results of a statistical analysis, such as the likelihood analysis used here. Our calculations indicate that the observations are somewhat more consistent with CHAZ$_{CRH}$, followed by MRI-LENS, Hi-TRACK, Hi-TempExt.

The uncertainty in CHAZ's projections of annual TC frequency comes from our design of the CHAZ model, but is broadly consistent with the level of broader understanding of TC frequency at present, including that derived from the latest high-resolution models shown here as well as other downscaling systems (Sobel et al., 2021). Other aspects of TC characteristics that could change with anthropogenic climate change have been also evaluated here, namely the forward motion and LMI95, are less dramatically uncertain, although our analyses show that one cannot rule out the role of natural variability. Still, the uncertainty regarding TC frequency introduces a large uncertainty into any assessment of overall TC risk, since any change of TC properties is only relevant to the extent that TCs actually occur.

The divergence between increasing and decreasing TC frequency scenarios becomes most pronounced in the latter part of the 21st century, but has some impact on the present and near future as well (Lee et al., 2020, 2022). In the situation when the change of frequency is subtle, changes in other TC properties may lead to differences in regional TC risk assessment.

How one views the situation must ultimately be based on one's particular application and the consequences of changing TC frequency. A priori, though, we argue that the most rational way to treat epistemic uncertainty is to consider all outcomes contained in the results to be possible. In terms of risk, it is useful to note that while in everyday language risk is associated with danger, harm, or negative consequences, in business and finance, risk simply refers to the uncertainty of potential outcomes, can be positive or negative. From that perspective the epistemic uncertainty present in TC climate change signal can only serve to increase risk.

Data Availability Statement

CHAZ (Lee et al., 2018) is an open-sourced model (Lee & Schechter, 2023). IBTrACS (Knapp et al., 2010) data are available at https://www.ncdc.noaa.gov/ibtracs/ (Knapp et al., 2022). The CAM5 data is available at https://portal.nersc.gov/c20c/ (C20C+ Detection and Attribution Sub-project, 2023; Stone et al., 2019) as the LBNLCAM5-1-2-025° model. HighResMIP TC information can be found at

https://catalogue.ceda.ac.uk/uuid/0b42715a7a804290afa9b7e31f5d7753 and https://catalogue.ceda.ac.uk/uuid/438268b75fed4f27988dc02f8a1d756d (Roberts, 2019). Underlying data for this publications are at https://doi.org/10.5281/zenodo.8417288 (Lee, 2023).

References

Baldwin, J. W., Lee, C.-Y., Walsh, B. J., Camargo, S. J., & Sobel, A. H. (2023). Vulnerability in a tropical cyclone risk model: Philippines case study. *Weather, Climate, and Society*, *15*, 503–523. https://doi.org/10.1175/WCAS-D-22-0049.1

Barnston, A. G., Li, S., Mason, S. J., DeWitt, D. G., Goddard, L., & Gong, X. (2010). Verification of the first 11 years of IRI's seasonal climate forecasts. *Journal of Applied Meteorology and Climatology*, *49*(3), 493–520. https://doi.org/10.1175/2009jamc2325.1

Bellomo, K., Murphy, L. N., Cane, M. A., Clement, A. C., & Polvani, L. M. (2018). Historical forcings as main drivers of the Atlantic multidecadal variability in the CESM large ensemble. *Climate Dynamics*, *50*(9), 3687–3698. https://doi.org/10.1007/s00382-017-3834-3

Beven, K. J., Aspinall, W. P., Bates, P. D., Borgomeo, E., Goda, K., Hall, J. W., et al. (2018). Epistemic uncertainties and natural hazard risk assessment—Part 2: What should constitute good practice? *Natural Hazards and Earth System Sciences*, *18*(10), 2769–2783. https://doi.org/10.5194/nhess-18-2769-2018

Bhatia, K. T., Vecchi, G., Murakami, H., Underwood, S., & Kossin, J. (2018). Projected response of tropical cyclone intensity and intensification in a global climate model. *Journal of Climate*, *31*(20), 8281–8303. https://doi.org/10.1175/jcli-d-17-0898.1

Bhatia, K. T., Vecchi, G. A., Knutson, T. R., Murakami, H., Kossin, J., Dixon, K. W., & Whitlock, C. E. (2019). Recent increases in tropical cyclone intensification rates. *Nature Communications*, *10*(1), 635. https://doi.org/10.1038/s41467-019-08471-z

Bister, M., & Emanuel, K. A. (1997). The genesis of hurricane Guillermo: TEXMEX analyses and a modeling study. *Monthly Weather Review*, *125*(10), 2662–2682. https://doi.org/10.1175/1520-0493(1997)125<2662:tgohgt>2.0.co;2

Blanusa, M. L., López-Zurita, C. J., & Rasp, S. (2022). The role of internal variability in global climate projections of extreme events. In *Physics—Atmospheric and oceanic physics*. Retrieved from http://arxiv.org/abs/2208.08275

Bloemendaal, N., Haigh, I. D., de Moel, H., Muis, S., Haarsma, R. J., & Aerts, J. C. J. H. (2020). Generation of a global synthetic tropical cyclone hazard dataset using STORM. *Scientific Data*, *7*(1), 40. https://doi.org/10.1038/s41597-020-0381-2

Bretherton, C. S., Peters, M. E., & Back, L. E. (2004). Relationships between water vapor path and precipitation over the tropical oceans. *Journal of Climate*, *17*(7), 1517–1528. https://doi.org/10.1175/1520-0442(2004)017<1517:rbwvpa>2.0.co;2

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, *11*(4), 995–1012. https://doi.org/10.5194/esd-11-995-2020

C20C+ Detection and Attribution Sub-project. (2023). [Dataset]. Retrieved from https://portal.nersc.gov/c20c/

Camargo, S. J., Giulivi, C., Sobel, A. H., Wing, A. A., Kim, D., Moon, Y., et al. (2020). Characteristics of model tropical cyclone climatology and the large-scale environment. *Journal of Climate*, *33*(11), 4463–4487. https://doi.org/10.1175/jcli-d-19-0500.1

Camargo, S. J., Tippett, M. K., Sobel, A. H., Vecchi, G. A., & Zhao, M. (2014). Testing the performance of tropical cyclone genesis indices in future climates using the HiRAM model. *Journal of Climate*, *27*(24), 9171–9196. https://doi.org/10.1175/jcli-d-13-00505.1

Chan, D., Vecchi, G. A., Yang, W., & Huybers, P. (2022). Improved simulation of 19th- and 20th-century North Atlantic hurricane frequency after correcting historical sea surface temperatures. *Science Advances*, *7*(26), eabg6931. https://doi.org/10.1126/sciadv.abg6931

Chan, J. C. (2008). Decadal variations of intense typhoon occurrence in the western North Pacific. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *464*(2089), 249–272. https://doi.org/10.1098/rspa.2007.0183

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., et al. (2019). Global mean climate and main patterns of variability in the CMCC-CM2 coupled model. *Journal of Advances in Modeling Earth Systems*, *11*(1), 185–209. https://doi.org/10.1029/2018ms001369

Clement, A., Bellomo, K., Murphy, L. N., Cane, M. A., Mauritsen, T., Radel, G., & Stevens, B. (2015). The Atlantic Multidecadal Oscillation without a role for ocean circulation. *Science*, *350*(6258), 320–324. https://doi.org/10.1126/science.aab3980

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., et al. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL Global Coupled Model CM3. *Journal of Climate*, *24*(13), 3484–3519. https://doi.org/10.1175/2011jcli3955.1

Emanuel, K. A. (1989). The finite-amplitude nature of tropical cyclogenesis. *Journal of the Atmospheric Sciences*, *46*(22), 3431–3456. https://doi.org/10.1175/1520-0469(1989)046<3431:tfanot>2.0.co;2

Emanuel, K. A. (2007). Environmental factors affecting tropical cyclone power dissipation. *Journal of Climate*, *20*(22), 5497–5509. https://doi.org/10.1175/2007jcli1571.1

Emanuel, K. A. (2008). The hurricane-climate connection. *Bulletin America Meteorology Social*, *89*(5), ES10–ES20. https://doi.org/10.1175/bams-89-5-emanuel

Emanuel, K. A. (2013). Downscaling CMIP5 climate models shows increased tropical cyclone activity over the 21st century. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(30), 12219–12224. https://doi.org/10.1073/pnas.1301293110

Emanuel, K. A. (2021). Atlantic tropical cyclones downscaled from climate reanalyses show increasing activity over past 150 years. *Nature Communications*, *12*(1), 7027. https://doi.org/10.1038/s41467-021-27364-8

Emanuel, K. A. (2022). Tropical cyclone seeds, transition probabilities, and genesis. *Journal of Climate*, *35*(11), 3557–3566. https://doi.org/10.1175/jcli-d-21-0922.1

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., et al. (2011). The community climate system model version 4. *Journal of Climate*, *24*(19), 4973–4991. https://doi.org/10.1175/2011jcli4083.1

Goldenberg, S. B., Landsea, C. W., Mestas-Nuñez, A. M., & Gray, W. M. (2001). The recent increase in Atlantic hurricane activity: Causes and implications. *Science*, *293*(5529), 474–479. https://doi.org/10.1126/science.1060040

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., et al. (2019). Max Planck institute Earth system model (MPI-ESM1.2) for the high-resolution model intercomparison project (HighResMIP). *Geoscientific Model Development*, *12*(7), 3241–3281. https://doi.org/10.5194/gmd-12-3241-2019

Haarsma, R. J., Acosta, M., Bakhshi, R., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., et al. (2020). HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR—Description, model computational performance and basic validation. *Geoscientific Model Development*, *13*(8), 3507–3527. https://doi.org/10.5194/gmd-13-3507-2020

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., et al. (2016). High resolution model intercomparison project (highresmip v1.0) for CMIP6. *Geoscientific Model Development*, *9*(11), 4185–4208. https://doi.org/10.5194/gmd-9-4185-2016

Hagen, A. B., & Landsea, C. W. (2012). On the classification of extreme Atlantic hurricanes utilizing mid-twentieth-century monitoring capabilities. *Journal of Climate*, *25*(13), 4461–4475. https://doi.org/10.1175/jcli-d-11-00420.1

Hassanzadeh, P., Lee, C.-Y., Nabizadeh, E., Camargo, S. J., Ma, D., & Yeung, L. Y. (2020). Effects of climate change on the movement of future landfalling Texas tropical cyclones. *Nature Communications*, *11*(1), 3319. https://doi.org/10.1038/s41467-020-17130-7

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108. https://doi.org/10.1175/2009bams2607.1

Hegerl, G., & Zwiers, F. (2011). Use of models in detection and attribution of climate change. *WIREs Climate Change*, *2*(4), 570–591. https://doi.org/10.1002/wcc.121

Hodges, K., Cobb, A., & Vidale, P. L. (2017). How well are tropical cyclones represented in reanalysis data sets? *Journal of Climate*, *30*(14), 5243–5264. https://doi.org/10.1175/jcli-d-16-0557.1

Intergovernmental Panel on Climate Change (IPCC). (2014). *Near-term climate change: Projections and predictability* (pp. 953–1028). Cambridge University Press.

Jones, C. D., Hughes, J. K., Bellouin, N., Hardiman, S. C., Jones, G. S., Knight, J., et al. (2011). The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geoscientific Model Development*, *4*(3), 543–570. https://doi.org/10.5194/gmd-4-543-2011

Klotzbach, P. J., & Gray, W. M. (2008). Multidecadal variability in North Atlantic tropical cyclone activity. *Journal of Climate*, *21*(15), 3929–3935. https://doi.org/10.1175/2008jcli2162.1

Knapp, K. R., Diamond, H. J., Kossin, J. P., Kruk, C. J. I., & Schreck, M. C. (2022). International best track archive for climate stewardship (IBTrACS) project, version 4 [Dataset]. NOAA National Centers for Environmental Information. https://doi.org/10.25921/82ty-9e16

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., & Neumann, C. J. (2010). The international best track archive for climate stewardship (IBTrACS). *Bulletin America Meteorology Social*, *91*(3), 363–376. https://doi.org/10.1175/2009bams2755.1

Knutson, T. R. (2017). *Climate science special report: A sustained assessment activity of the U.S. Global change research program*. In D. Wuebbles, D. Fahey, K. Hibbard, D. Dokken, B. Stewart, & T. Maycock (Eds.), (pp. 652–663). U.S. Global Change Research Program.

Knutson, T. R., Camargo, S., Chan, J., Emanuel, K., Ho, C.-H., Kossin, J., et al. (2020a). Tropical cyclones and climate change assessment: Part I. Detection and attribution. *Bulletin America Meteorology Social*, *100*(10), 1987–2007. https://doi.org/10.1175/bams-d-18-0189.1

Knutson, T. R., Camargo, S. J., Chan, J. C. L., Emanuel, K., Ho, C.-H., Kossin, J., et al. (2020b). Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bulletin America Meteorology Social*, *101*(3), E303–E322. https://doi.org/10.1175/bams-d-18-0194.1

Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K. A., Holland, G., Landsea, C., et al. (2010). Tropical cyclones and climate change. *Nature Geoscience*, *3*, 157–163. https://doi.org/10.1038/ngeo779

Knutson, T. R., Sirutis, J. J., Garner, S. T., Held, I. M., & Tuleya, R. E. (2007). Simulation of the recent multidecadal increase of Atlantic hurricane activity using an 18-km-grid regional model. *Bulletin of the American Meteorological Society*, *88*(10), 1549–1565. https://doi.org/10.1175/bams-88-10-1549

Kossin, J. P. (2018). A global slowdown of tropical-cyclone translation speed. *Nature*, *558*(7708), 104–107. https://doi.org/10.1038/s41586-018-0158-3

Kossin, J. P., Olander, T. L., & Knapp, K. R. (2013). Trend analysis with a new global record of tropical cyclone intensity. *Journal of Climate*, *26*(24), 9960–9976. https://doi.org/10.1175/jcli-d-13-00262.1

Landsea, C. W., Vecchi, G. A., Bengtsson, L., & Knutson, T. R. (2010). Impact of duration thresholds on Atlantic tropical cyclone counts. *Journal of Climate*, *23*(10), 2508–2519. https://doi.org/10.1175/2009jcli3034.1

Lee, C.-Y. (2023). Data for Lee, Sobel, Tippett, Camargo, Wüst, Wehner, Murakami (2023) Earth's future [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.8417288

Lee, C.-Y., Camargo, S. J., Sobel, A. H., & Tippett, M. K. (2020). Statistical-dynamical downscaling projections of tropical cyclone activity in a warming climate: Two diverging genesis scenarios. *Journal of Climate*, *33*(11), 4815–4834. https://doi.org/10.1175/jcli-d-19-0452.1

Lee, C.-Y., & Schechter, E. (2023). CHAZ: The Columbia HAZard model [Software]. GitHub. Retrieved from https://github.com/cl3225/CHAZ

Lee, C.-Y., Sobel, A. H., Camargo, S. J., Tippett, M. K., & Yang, Q. (2022). New York State hurricane hazard: History and future projections. *Journal of Applied Meteorology and Climatology*, *61*(6), 61–629. https://doi.org/10.1175/jamc-d-21-0173.1

Lee, C.-Y., Tippett, M. K., Camargo, S. J., & Sobel, A. H. (2015). Probabilistic multiple linear regression modeling for tropical cyclone intensity. *Monthly Weather Review*, *143*(3), 933–954. https://doi.org/10.1175/mwr-d-14-00171.1

Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2016). Autoregressive modeling for tropical cyclone intensity climatology. *Journal of Climate*, *29*(21), 7815–7830. https://doi.org/10.1175/jcli-d-15-0909.1

Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2018). An environmentally forced tropical cyclone hazard model. *Journal of Advances in Modeling Earth Systems*, *10*(1), 223–241. https://doi.org/10.1002/2017ms001186

Lloyd, E. A., & Oreskes, N. (2018). Climate change attribution: When is it appropriate to accept new methods? *Earth's Future*, *6*(3), 311–325. https://doi.org/10.1002/2017ef000665

Mann, M. E., & Emanuel, K. A. (2006). Atlantic hurricane trends linked to climate change. *Eos, Transactions American Geophysical Union*, *87*(24), 233–241. https://doi.org/10.1029/2006eo240001

Meiler, S., Vogt, T., Bloemendaal, N., Ciullo, A., Lee, C.-Y., Camargo, S. J., et al. (2022). Intercomparison of regional loss estimates from global synthetic tropical cyclone models. *Nature Communications*, *13*(1), 6156. https://doi.org/10.1038/s41467-022-33918-1

Mizuta, R., Murata, A., Ishii, M., Shiogama, H., Hibino, K., Mori, N., et al. (2017). Over 5,000 years of ensemble future climate simulations by 60-km global and 20-km regional atmospheric models. *Bulletin America Meteorology Social*, *98*(7), 1383–1398. https://doi.org/10.1175/bams-d-16-0099.1

Moon, Y., Kim, D., Wing, A. A., Camargo, S. J., Zhao, M., Leung, L. R., et al. (2022). An evaluation of tropical cyclone rainfall structures in the HighResMIP simulations against satellite observations. *Journal of Climate*, *35*(22), 1–60. https://doi.org/10.1175/jcli-d-21-0564.1

Murakami, H., Mizuta, R., & Shindo, E. (2012). Future changes in tropical cyclone activity projected by multi-physics and multi-SST ensemble experiments using the 60-km-mesh MRI-AGCM. *Climate Dynamics*, *39*(9), 2569–2584. https://doi.org/10.1007/s00382-011-1223-x

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., et al. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Roberts, M. J. (2019). CMIP6 HighResMIP: Tropical storm tracks [Dataset]. Centre for Environmental Data Analysis. Retrieved from http://catalogue.ceda.ac.uk/uuid/e82a62d926d7448696a2b60c1925f811

Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., et al. (2019). Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geoscientific Model Development*, *12*(12), 4999–5028. https://doi.org/10.5194/gmd-12-4999-2019

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vannière, B., et al. (2020b). Projected future changes in tropical cyclones using the CMIP6 HighResMIP multimodel ensemble. *Geophysical Research Letters*, *47*(14), e2020GL088662. https://doi.org/10.1029/2020gl088662

Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., et al. (2020a). Impact of model resolution on tropical cyclone simulation using the HighResMIP-PRIMAVERA multimodel ensemble. *Journal of Climate*, *33*(7), 2557–2583. https://doi.org/10.1175/jcli-d-19-0639.1

Rousseau-Rizzi, R., & Emanuel, K. (2020). Estimating the causes of past Atlantic tropical cyclone multidecadal variability. In *AGU Fall Meeting*.

Schreck, C. J., Knapp, K. K., & Kossin, J. P. (2014). The impact of best track discrepancies on global tropical cyclone climatologies using IBTrACS. *Monthly Weather Review*, *142*(10), 3881–3899. https://doi.org/10.1175/mwr-d-14-00021.1

Shepherd, T. G. (2019). Storyline approach to the construction of regional climate change information. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *475*(2225), 20190013. https://doi.org/10.1098/rspa.2019.0013

Sobel, A. H., Camargo, S. J., & Previdi, M. (2019). Aerosol versus greenhouse gas effects on tropical cyclone potential intensity and the hydrologic cycle. *Journal of Climate*, *32*(17), 5511–5527. https://doi.org/10.1175/jcli-d-18-0357.1

Sobel, A. H., Lee, C.-Y., Camargo, S. J., Mandli, K. T., Emanuel, K. A., Mukhopadhyay, P., & Mahakur, M. (2019). Tropical cyclone hazard to Mumbai in the recent historical climate. *Monthly Weather Review*, *147*(7), 2355–2366. https://doi.org/10.1175/mwr-d-18-0419.1

Sobel, A. H., Wing, A. A., Camargo, S. J., Patricola, C. M., Vecchi, G. A., Lee, C.-Y., & Tippett, M. K. (2021). Tropical cyclone frequency. *Earth's Future*, *9*(12), e2021EF002275. https://doi.org/10.1029/2021ef002275

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., et al. (2007). *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Tech. Rep.). Cambridge University Press.

Stone, D. A., Christidis, N., Folland, C., Perkins-Kirkpatrick, S., Perlwitz, J., Shiogama, H., et al. (2019). Experiment design of the international CLIVAR C20C+ detection and attribution project. *Weather and Climate Extremes*, *24*, 100206. https://doi.org/10.1016/j.wace.2019.100206

Sugi, M., Yamada, Y., Yoshida, K., Mizuta, R., Nakano, M., Kodama, C., & Satoh, M. (2020). Future changes in the global frequency of tropical cyclone seeds. *SOLA*, *16*(0), 70–74. https://doi.org/10.2151/sola.2020-012

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin America Meteorology Social*, *93*(4), 485–498. https://doi.org/10.1175/bams-d-11-00094.1

Ting, M., Kushnir, Y., Seager, R., & Li, C. (2011). Robust features of Atlantic multi-decadal variability and its climate impacts. *Geophysical Research Letters*, *38*(17), L17705. https://doi.org/10.1029/2011gl048712

Tippett, M., Camargo, S. J., & Sobel, A. H. (2011). A Possion regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis. *Journal of Climate*, *21*(9), 2335–2357. https://doi.org/10.1175/2010jcli3811.1

Toumi, R., & Restell, L. (2014). Catastrophe modelling and climate change. (Tech. Rep.). Lloyd's.

Ullrich, P. A., & Zarzycki, C. M. (2017). TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, *10*(3), 1069–1090. https://doi.org/10.5194/gmd-10-1069-2017

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., & Reed, K. A. (2021). TempestExtremes v2.1: A community framework for feature detection, tracking and analysis in large datasets. *Geoscientific Model Development*, *2021*(8), 1–37. https://doi.org/10.5194/gmd-14-5023-2021

Vecchi, G. A., Delworth, T. L., Murakami, H., Underwood, S. D., Wittenberg, A. T., Zeng, F., et al. (2019). Tropical cyclone sensitivities to CO_2 doubling: Roles of atmospheric resolution, synoptic variability and background climate changes. *Climate Dynamics*, *53*(9), 5999–6033. https://doi.org/10.1007/s00382-019-04913-y

Vecchi, G. A., & Knutson, T. R. (2008). On estimates of historical North Atlantic tropical cyclone activity. *Journal of Climate*, *21*(14), 3580–3600. https://doi.org/10.1175/2008jcli2178.1

Vecchi, G. A., Landsea, C., Zhang, W., Villarini, G., & Knutson, T. (2021). Changes in Atlantic major hurricane frequency since the late-19th century. *Nature Communications*, *12*(1), 4054. https://doi.org/10.1038/s41467-021-24268-5

Villarini, G., Vecchi, G. A., Knutson, T. R., & Smith, J. A. (2011). Is the recorded increase in short-duration North Atlantic tropical storms spurious? *Journal of Geophysical Research*, *116*(D10), D10114. https://doi.org/10.1029/2010jd015493

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al. (2019). Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2177–2213. https://doi.org/10.1029/2019ms001683

Wang, X., Wang, C., Zhang, L., & Wang, X. (2015). Multidecadal variability of tropical cyclone rapid intensification in the western North Pacific. *Journal of Climate*, *28*(9), 3806–3820. https://doi.org/10.1175/jcli-d-14-00400.1

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., et al. (2010). Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *Journal of Climate*, *23*, 6312–6335. https://doi.org/10.1175/2010jcli3679.1

Wehner, M. F., Prabhat, Reed, K. A., Stone, D., Collins, W. D., & Bacmeister, J. (2015). Resolution dependence of future tropical cyclone projections of CAM5.1 in the U.S. CLIVAR hurricane working group idealized configurations. *Journal of Climate*, *28*(10), 3905–3925. https://doi.org/10.1175/jcli-d-14-00311.1

Wehner, M. F., Reed, K. A., Li, F., Prabhat, Bacmeister, J., Chen, C.-T., & Jablonowski, C. (2014). The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. *Journal of Advances in Modeling Earth Systems*, *6*(4), 980–997. https://doi.org/10.1002/2013ms000276

Wehner, M. F., Reed, K. A., Loring, B., Stone, D., & Krishnan, H. (2018). Changes in tropical cyclones under stabilized 1.5 and 2.0°C global warming scenarios as simulated by the Community Atmospheric Model under the HAPPI protocols. *Earth System Dynamics*, *9*(187–195), 187–195. https://doi.org/10.5194/esd-9-187-2018

Yan, X., Zhang, R., & Knutson, T. R. (2017). The role of Atlantic overturning circulation in the recent decline of Atlantic major hurricane frequency. *Nature Communications*, *8*(1), 1695. https://doi.org/10.1038/s41467-017-01377-8

Yoshida, K., Sugi, M., Mizuta, R., Murakami, H., & Ishii, M. (2017). Future changes in tropical cyclone activity in high-resolution large-ensemble simulations. *Geophysical Research Letters*, *44*(19), 9910–9917. https://doi.org/10.1002/2017gl075058

Zanchettin, D., Timmreck, C., Graf, H.-F., Rubino, A., Lorenz, S., Lohmann, K., et al. (2012). Bi–decadal variability excited in the coupled ocean–atmosphere system by strong tropical volcanic eruptions. *Climate Dynamics*, *39*(1–2), 419–444. https://doi.org/10.1007/s00382-011-1167-1

Zarzycki, C. M., & Ullrich, P. A. (2017). Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophysical Research Letters*, *44*(2), 1141–1149. https://doi.org/10.1002/2016gl071606