Climate change signal in Atlantic tropical cyclones today and near future

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

[1]Columbia University
[2]Marie Tharp Lamont Research Professor, Ocean and Climate Physics Division, Lamont-Doherty Earth Observatory, Columbia University
[3]Swiss Re
[4]Lawrence Berkeley National Laboratory (DOE)
[5]GFDL

February 27, 2023

Abstract

This manuscript discusses the challenges in detecting and attributing recently observed trends in the Atlantic hurricanes and the epistemic uncertainty we face in assessing future hurricane risk. Data used here include 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 hurricane activity: the upward trend and multi-decadal oscillation of the annual frequency, the increase in storm wind intensity, and the downward trend in the forward speed. Some datasets suggest that these trends and oscillation are forced while others suggest that they can be explained by natural variability. Future 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 the future hurricane risk assessment. Here, we investigate the reduction of epistemic uncertainties on annual frequency through a statistical practice – likelihood analysis. We find that historical observations are more consistent with the simulations with increasing frequency but we are not able to rule out other possibilities. We argue that the most rational way to treat epistemic uncertainty is to consider all outcomes contained in the results. In the context of hurricane risk assessment, since the results contain possible outcomes in which hurricane risk is increasing, this view implies that the risk is increasing.

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], 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

Key Points:

- Changes in the Atlantic hurricane 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
- Based on our results, it is more likely that the risk of hurricanes is increasing than that it is decreasing, though not by a large margin

Corresponding author: Chia-Ying Lee, `cl3225@columbia.edu`

Abstract

This manuscript discusses the challenges in detecting and attributing recently observed trends in the Atlantic hurricanes and the epistemic uncertainty we face in assessing future hurricane risk. Data used here include 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 hurricane activity: the upward trend and multi-decadal oscillation of the annual frequency, the increase in storm wind intensity, and the downward trend in the forward speed. Some datasets suggest that these trends and oscillation are forced while others suggest that they can be explained by natural variability. Future 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 the future hurricane risk assessment. Here, we investigate the reduction of epistemic uncertainties on annual frequency through a statistical practice – likelihood analysis. We find that historical observations are more consistent with the simulations with increasing frequency but we are not able to rule out other possibilities. We argue that the most rational way to treat epistemic uncertainty is to consider all outcomes contained in the results. In the context of hurricane risk assessment, since the results contain possible outcomes in which hurricane 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 hurricanes. We looked at three aspects of these storms: the number of hurricanes 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 hurricane risk requires us to understand how the number of hurricanes 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 hurricanes 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 (Schreck et al., 2014; Emanuel, 2021; D. Chan et al., 2022). 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.

Accounting for climate change is likely to require a greater use of physics than is historically typical in catastrophe models (Toumi & Restell, 2014; Emanuel, 2008). 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., Klotzbach & Gray, 2008; J. C. Chan, 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 cyclones (TCs) 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 processes, 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), near future (up to 2040 or 2050), and pre-industrial control period. Simulations from pre-industrial control period contain no anthropogenic forcing signal 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 (tropical cyclone) 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 organically (Yoshida et al., 2017; Wehner et al., 2014; Roberts et al., 2020).

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 (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 aerosols over the Atlantic since 1980s has been argued to be a dominant cause of the increasing TC activity in late 20$^{\text{th}}$ century (Mann & Emanuel, 2006; Sobel, Camargo, & Previdi, 2019; Rousseau-Rizzi & Emanuel, 2020). On the other hand, several measures of Atlantic TC activity, including the major hurricane (TCs with LMI $\geq 93\,\text{kt}$) frequency (Goldenberg et al., 2001), are highly correlated to the the Atlantic Multi-decadal Oscillation (AMO) or Atlantic multidecadal variability (AMV), a low-frequency mode of variability identified by the average sea surface temperature anomalies in the North Atlantic basin, typically over 0-80$^{\text{o}}$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 (Clement et al., 2015; Bellomo et al., 2018). 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 the Atlantic TCs with warming (e.g., Knutson et al., 2010, and others), with a few exceptions (Emanuel, 2013; Bhatia et al., 2018; Vecchi et al., 2019).

The second objective is to compare historical simulations with observations to understand which modeling dataset 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 the global tropical cyclone genesis (i.e. Sobel et al., 2021). Lastly, we will assess hurricane risk over a set of selected 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 datasets

2.1.1 Observations

For reference, we use best-track data from National Hurricane Center obtained via International Best Track Archive for Climate Stewardship v04r00 IBTrACS (Knapp et al., 2010). We use 6-hourly storm positions (in longitude and latitude) and maximum wind speeds (kt) from 1951 to 2020. Storm forward speed is derived from the position data. We use only storms whose lifetime maximum intensity (LMI) reaches tropical storm (TS) strength, 34 kt. Hurricanes are referred to storms with LMI of at least 64 kt.

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 (tropical cyclone) 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. There is no feedback of 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 to 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 tropical cyclone 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) and Lee et al. (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. 2022). Meiler et al. (2022) found that losses estimated from CHAZ are comparable to those estimated using comparable academic tropical cyclone 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 (NCAR) Community Climate System Model 4 (CCSM4) (Gent et al., 2011), the Geophysical Fluid Dynamics Laboratory Climate Model version 3 (GFDL-CM3) (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 (MIROC5) (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 (Camargo et al., 2014). 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 thermodynamic inhibition of TC formation in a warming climate (Emanuel, 1989, 2022), but the gaps in our understanding of the relationship between climate and tropical cyclone 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 tropical cyclone 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 $\mathrm{CHAZ}_{\mathrm{CRH}}$ and $\mathrm{CHAZ}_{\mathrm{SD}}$.

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, i.e., tropical cyclone-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 was discussed in Yoshida et al. (2017). The second one is the 25-km High-Resolution Community Atmospheric Model version 5 simulations, CAM5 (Wehner et al., 2014, 2015). Next, we use storms tracked in the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016) High Resolution Model Intercomparison Project (HighResMIP) (Haarsma et al., 2016). Following Roberts et al. (2020) and Roberts et al. (2020), 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; Zarzycki & Ullrich, 2017; Ullrich et al., 2021), and both event sets are used here. 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 $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$, 1950-2010 for MRI-LENS, 1996-2016 for CAM5, and 1951-2014 for the two HighResMIP datasets. In the future period, $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$ contain storms from 2006-2040 under Representative Concentration Pathway 8.5 (rcp8.5) while HighResMIP storms are from 2015-2050 under Shared Socioeconomic Pathways5-85 (ssp585). Both are high-emission scenarios with an additional radiative forcing of 8.5 W $\mathrm{m^{-2}}$ by the year 2100 (Riahi et al., 2017) in ssp585 which considers a fossil-fueled development. 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 datasets as all years are equivalent. The MIR-LENS and CAM5 piC simulations are exceptions. In MRI-LENS and CAM5, the observed SST information is given in both historical and piC simulations as a lower boundary, but the long-term trend is removed in the piC simulations. In other words, MIR-LENS and CAM5 piC simulations still contain observed variation. 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 that only the linear trend is anthropogenic forced, not the variability, which, as we will discussed 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 4000 considering intensity ensemble), as is possible because the CHAZ intensity module has a stochastic component. Hi-TRACK has 7 members (5 global climate models and two of them have 2 ensemble members) and Hi-TempExt has 6 (4 global climate models and two of them have 2 ensemble members). MRI-LENS has 100 ensemble members while CAM5 has 5. The data properties are listed in Table 1.

2.3 Frequency adjustment

There are biases in 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-intensity biases, in general, as the grid spacings of these models are too coarse to capture the full range of observed hurricane strengths (e.g., Yoshida et al., 2017; Moon et al., 2022, and others). Here we address only the frequency biases. Specifically, we derive an adjustment by comparing the basin-wide annual TC frequency of models' historical simulations to that of the observations from the same period. The same adjustment will then be 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) and Lee et al. (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 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:

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

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 dataset are from its respective historical simulation. Observations are calculated from their respective historical periods. 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 (in Section 5), we apply a multiplicative factor to ensure the annual frequency at storm with intensity greater than 40 kt in these data sets are consistent to observations, which is the same as the bias-correction approach used in (Lee et al., 2022).

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 rcp and hist 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 (Clement et al., 2015; Bellomo et al., 2018). 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 unites of the variable's unit per year (yr^{-1}) and per year square (yr^{-2}), respectively.

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). Figure 1a and b show the ensemble means of the time series of Atlantic hurricane frequency, i.e., the averaged total number of storms in the basin each year whose maximum sustained winds exceed 34 kt from each dataset. The small wiggles may be sampling variability. Figures 1c and d show the ensemble spread. By construction, the time-mean annual frequency for each dataset over its respective historical period will be identical to observations after the frequency adjustment (Eq. (1)). The original annual frequency of each dataset 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 $\mathrm{CHAZ_{SD}}$ (orange thick line) and $\mathrm{CHAZ_{CRH}}$ (blue thick line) results begin to diverge, with $\mathrm{CHAZ_{SD}}$ showing a decreasing TC frequency and $\mathrm{CHAZ_{CRH}}$ 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 HiResMIP datasets show no considerable trend in the historical period but a sharp dip after 2030. The ssp585 scenario in HiResMIP starts after 2015, though. Hi-TRACK's annual TC frequency climbs up by 2040. Roberts et al. (2020) 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 to the mean trend.

Figures 1b and 1d show analogous results for piC simulations. Note that the years in the x-axis are not real; 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 the Figure 1b, $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$ results do not diverge. There is no dip in the Hi-TRACK or Hi-TempExt. Clearly, the separation between $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$ and the dip in the two HighResMIP datasets in Figure 1a represent forced responses.

Next we conduct the trend analyses of the annual TC frequency in Figure 1 using second-order Legendre polynomials fits (Eq. (2)). As an example, Fig. 2a shows the analysis using the $\mathrm{CHAZ_{CRH}}$ simulations and the observations. The observed fit (dashed black line) has an upward trend of 0.085 storm year^{-2} and a positive curvature of 0.005 storm year^{-2} (shown as the black line in Figs. 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 (Clement et al., 2015; Bellomo et al., 2018). In Fig. 2a, the polynomial fits of $\mathrm{CHAZ_{CRH}}$ simulations from historical only (light blue dashed line) and from historical to future (dark blue dashed line) both show an small upward curve 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 dataset are also shown in Figures 2b-c. The observed linear trend are 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 lies within the 25-75 percentile ensemble ranges of piC simulations from $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$, and MRI-LENS. It is outside of the ensemble ranges from two HighResMIP datasets which have quadratic terms close to zero. The observed linear trend is at top 25 percentile (75-100 percentile) of the hist simulations of $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$, and is marginally included in the simulations of MRI-LENS; the observed quadratic term is within the 25-75 percentile range the $\mathrm{CHAZ_{CRH}}$ and MRI-LENS, and is at top 25 percentile in $\mathrm{CHAZ_{SD}}$. Only the fit linear trend derived from historical + future simulations of the $\mathrm{CHAZ_{CRH}}$ include the observed value. For the quadratic trend, the fit parameter derived from $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$ include the observed values but they are at top and bottom 25 percentile range, respectively. (We do not use any warming simulations from CAM5 and MRI-LENS.)

Generally speaking, the polynomial fit analysis suggests that, first, $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$ and MRI-LENS are better in capturing the observed trend and multi-decadal variability as their historical spread covers the observed values. However, CAM5 has only 10 years of data with 5 ensemble members and while Hi-TRACK and Hi-TempExt have only, respectively, 7 and 6 ensemble members. These three datasets may be under-sampled. Second, the observed linear trend is outside the spread of $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$ and MRI-LENS' piC simulations but within the spread of these models' hist simulations, indicating that anthropogenic forcing is necessary to capture the upward trend in the past decades. On the other hand, we can not rule out the possibility of the recent upward curvature trend is within the range of natural variability. Although the MRI-LENS' piC simulations is forced with the observed SST (with long-term trend removed) which results in the upward curvature term right on top of observed values in Figure 2c. Simulations from $\mathrm{CHAZ_{CRH}}$ suggest that that anthropogenic forcing helps to capture the upward curvature trend. Third, when considering the future period as well, the mean of $\mathrm{CHAZ_{CRH}}$ shows an upward trend, the mean of $\mathrm{CHAZ_{SD}}$ shows a downward trend, while the mean of the two HighResMIP simulations are close to zero. However, we have low confidence in the projections as they include zero. Thus, we can not say for sure that the positive linear and quadratic terms will continue into the future.

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

3.2 Intensity and storm motion

Figure 2d shows the fit parameters of Atlantic TC lifetime maximum intensity (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 linear term only. There is an upward trend in the observations, meaning that the extreme tail of observed intensity has increased with time, consistent with previous studies (e.g., Knutson et al., 2020a, and others). The positive linear trend is captured by the ensemble spreads of two CHAZ datasets and those of MRI-LENS and CAM5 at both piC and historical periods. It is outside of the ensemble spread of all simulations from from Hi-TRACK and Hi-TempExt. Thus, at least from $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$, MRI-LENS, and CAM5, we can not rule out that the recent upward trend in the LMI95 is due to natural variability. When looking into the future, only the means of $\mathrm{CHAZ_{CRH}}$ is positive and the means of $\mathrm{CHAZ_{SD}}$, 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, indicating, again, low-confidence in the projected changes.

Figure 2e shows the analysis for translation speed. Consistent with (Kossin, 2018), the observations show a clear downward trend in the storm motion. This trend is within ensemble spread in all periods, including piC, for all models, except the simulations from Hi-TempExt. However, the mean and the 25-75 percentile ensemble spreads in these datasets move toward different directions from piC to hist to hist +future periods. The Hi-Track and MRL-LENS hist simulations show upward trends in the storm motion and this upward trends continues in to the future. The differences in mean and 25-75 percentile ensemble spreads from $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$ from these three period are small. The piC and hist simulations from CAM5 shows that anthropogenic forcing may lead to a strong downward trend in storm motion but again CAM5 simulations are shorter than do the data from 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 motion speed for TCs affecting Texas (Hassanzadeh et al., 2020) and an 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 trend in historical and historical + future vary from one dataset to another. This is especially true for the TC frequency projections between $\mathrm{CHAZ}_{\mathrm{CRH}}$ and $\mathrm{CHAZ}_{\mathrm{SD}}$, 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, multi-model ensemble mean is a common approach to obtain the consensus from multiple global climate models. However, such approach is only adequate when the ensemble spread represents variations that can be considered random, as might be the case with typical aleatoric uncertainties. The divergent scenarios in the frequency projections are a consequence of the epistemic uncertainty due to the lack of a satisfactory scientific understanding of tropical cyclone frequency (Sobel et al., 2021; Emanuel, 2022) 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 in 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 dataset by fitting the model annual TC frequency to a Poisson regression. We do so for all simulations with data throughout 2021 (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, i.e., 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 $\mathrm{CHAZ}_{\mathrm{CRH}}$ simulations will generate 29 TCs as observed in 2005 is 0.08%, which is based on a Poisson distribution with $\lambda_{2005} = 15.7$. 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 larger than 0, the observations are more consistent with the simulations with anthropogenic forcing than with the piC simulations and vice versa.

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 $\mathrm{CHAZ}_{\mathrm{CRH}}$, MRI-LENS, and Hi-TempExt. The annual likelihood ratios from these three datasets 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 datasets when compared to in piC (Figure 2a). As λ_t increases with time, the distributions from these three datasets 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, $\mathrm{CHAZ}_{\mathrm{SD}}$ 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 $\mathrm{CHAZ}_{\mathrm{SD}}$. 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 we consider the likelihood over the whole observational period, we calculate the average of the likelihood, i.e., the roots of $\prod_{2021}^{1950} L_t$ from all five datasets. Between $\mathrm{CHAZ}_{\mathrm{CRH}}$ and $\mathrm{CHAZ}_{\mathrm{SD}}$, observations are more consistent with $\mathrm{CHAZ}_{\mathrm{CRH}}$, which has an averaged likelihood of 5.24%, than to $\mathrm{CHAZ}_{\mathrm{SD}}$ which has the averaged likelihood of

3.46%. Among the five datasets, $\mathrm{CHAZ_{CRH}}$ has highest likelihood, followed by Hi-TempExt (5.13%), MRI-LENS (5.1%), Hi-TRACK (5.04%), and CAM5 (3.6%).

The basin-wide frequency adjustment (Eq. (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 datasets 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 $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$,and MRI-LENS than in their respective piC simulations due to the large bias in these piC simulations. Without basin-wide TC frequency adjustment, Hi-TRACK has the greatest averaged likelihood, followed by CAM5, $\mathrm{CHAZ_{CRH}}$, Hi-TempExt, $\mathrm{CHAZ_{SD}}$, and MRI-LENS. MRI-LENS has the lowest likelihood because of its low bias and zero storms in some years.

5 Climate change and regional hurricane risk at three line gates

Now we compute regional hurricane risk, from hazard perspective only, represented by return periods of storms of given wind intensities passing through pre-defined gates, derived using simulations from historical and future periods. We use simulations from $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$, Hi-TRACK, and Hi-TempExt. The three line gates used here (black lines in Figures 4a–c) are the main development region (MDR) gate which can be thought of as delineating Atlantic TC hazard in a general sense – how many storms form, and at what intensity and move from the MDR toward the US and Caribbean Islands; the GoM gate which records TC activity for those 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 $\mathrm{CHAZ_{CRH}}$ with basin-wide and regional frequency adjustments (Eq. (1)). While the basin-wide frequency adjustment (dashed lines) yields a TC frequency close to observations at the GoM gate, $\mathrm{CHAZ_{CRH}}$ still overestimates storm activity at the MDR gate and underestimates storm activity at the NE gate. The regional frequency adjustment shifts the simulated return period curves (solid line, local adjustment) by matching the return periods at 40 kt to the values derived from observations (see Section 2.3 for details). In terms of the shape of the return period curve, as well as the return periods at high intensities, $\mathrm{CHAZ_{CRH}}$ 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–i show the return period curves derived from the four datasets 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 (Roberts et al., 2020; Moon et al., 2022). The return period curves of $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$ hist simulations are close to each other. The differences between simulations from historical period and those from historical and future periods, i.e., the differences between the dashed and solid lines, are small for the two CHAZ datasets in Figures 4g–i. 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-i. As expected, there is an overall increase in the storm frequency at all thresholds from historical to future periods for $\mathrm{CHAZ_{CRH}}$ while there is an overall decrease for $\mathrm{CHAZ_{SD}}$, consistent with the results in Figures 1 and 2a. The percentage changes are larger at higher intensity thresholds in the $\mathrm{CHAZ_{CRH}}$ but they are of similar or smaller magnitude throughout the Saffir-Simpson categories in the $\mathrm{CHAZ_{SD}}$. This is probably due to the increase in storm intensity as climate warms in $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$.

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+ ($\leq 64\,\mathrm{kt}$) 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+ storms, respectively. Hi-TempExt shows a 68% decrease and 16% increase of Category 1+ storms 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+ storms. 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 hurricane 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 $\mathrm{CHAZ_{CRH}}$ simulations and decreases in $\mathrm{CHAZ_{SD}}$, a difference that hangs on a subtle modeling choice (saturation deficit vs. relative humidity as a predictor of genesis). Changes in the high-resolution global climate model simulations are smaller, but again their direction depends on which global climate 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 dataset 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 $\mathrm{CHAZ_{CRH}}$, followed by Hi-TempExt, MRI-LENS, Hi-TRACK. However, the likelihood values among these four datasets are close to each other, so it would not be justified to draw a definitive conclusion from this analysis as to which model is most correct.

The epistemic uncertainty in CHAZ's projections on annual TC frequency comes from our design of the CHAZ model, but the conclusion is that our results are 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 attitude towards risk and the consequences of being wrong in either direction. 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 the present context, since the results contain possible outcomes in which TC risk — as estimated from a single model or subset of the entire multi-model ensemble — is increasing, that in itself means we should regard TC risk as increasing, at the highest level of understanding in which all available information is considered, even though there are other possible outcomes in which it is decreasing.

Open Research Section

CHAZ is an open-sourced model (`https://github.com/cl3225/CHAZ`). IBTrACS data are available at (`https://www.ncdc.noaa.gov/ibtracs/`). Information for CMIP5 data can be found at `https://pcmdi.llnl.gov/mips/cmip5/` and HighResMIP tropical cyclone information can be found at (`http://catalogue.ceda.ac.uk/uuid/e82a62d926d7448696a2b60c1925f81` Underlying data for this publications are at (`https://github.com/cl3225/Lee_etal_2023EarthsFuture`).

Acknowledgments

Lee, Sobel, Camargo and Tippett were supported by Swiss Re MGMT LTD (CU18-3118). Wehner was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE340AC02-05CH11231 under the Regional and Global Model Analysis (RGMA) program.

References

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. *J Appl Meteorol Climatol.*, *49*(3), 493–520.

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.

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. *J. Climate*, *31*, 8281–8303.

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. *Nat. Commun.*, *10*(1), 635.

Bister, M., & Emanuel, K. A. (1997). The genesis of hurricane Guillermo: TEXMEX analyses and a modeling study. *Mon. Wea. Rev.*, *125*(10), 2662–2682.

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.

Bretherton, C. S., Peters, M. E., & Back, L. E. (2004). Relationships between water vapor path and precipitation over the tropical oceans. *J. Climate*, *17*, 1517-1528.

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.

Table 1. Data Characteristics

data	global climate models	resolution	ens	period	annual frequency
$CHAZ_{CRH/SD}$	HadGEM2_ES	N/A	100	1951–2005;2006–2040; piC	8.8 /15.9
	CCSM4				11/16.1
	GFDL_CM3				16.5/19.1
	MPI_ESM_MR				29.3/39.4
	MIROC5				11.9/18.3
MRI-LENS	MRI–AGCM3.2H	60 km	100	1950–2010; piC	2.3
CAM5	CAM5	28km	5	1996–2005;piC	10.9
Hi-TRACK	CMCC-CM2-VHR4 (r1i1p1f1)	25 km	7	1951-2014;2015-2040;piC	5.0
	CNRM-CM6-1-HR(r1i1p1f2)	50 km			21.0
	EC-Earth3P-HR (r1i1p2f1)	50 km			6.8
	EC-Earth3P-HR (r2i1p2f1)	50 km			6.5
	HadGEM3-GC31-HH (r1i1p1f1)	50 km			21.5
	HadGEM3-GC31-HM (r1i1p1f1)	50 km			19.7
	MPI-ESM1-2-XR (r1i1p1f1)	50 km			4.5
Hi-TempExt	CNRM-CM6-1-HR(r1i1p1f2)	50 km	6	1951-2014;2015-2040;piC	13.4
	EC-Earth3P-HR (r1i1p2f1)	50 km			2
	EC-Earth3P-HR (r2i1p2f1)	50 km			2
	HadGEM3-GC31-HH (r1i1p1f1)	50 km			13.3
	HadGEM3-GC31-HM (r1i1p1f1)	50 km			12.4
	MPI-ESM1-2-XR (r1i1p1f1)	50 km			0.63

Camargo, S. J., Giulivi, C., Sobel, A. H., Wing, A. A., Kim, D., Moon, Y., … Zhao, M. (2020). Characteristics of model tropical cyclone climatology and the large-scale environment. *J. Climate*, *33*, 4463–4487.

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. *J. Climate*, *27*, 9171-9196.

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).

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.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., … Navarra, A. (2019). Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model. *J. Adv. Model. Earth Syst.*, *11*(1), 185–209.

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*, 320–324.

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., … Zeng, F. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL Global Coupled Model CM3. *J. Climate*, *24*, 3484–3519.

Emanuel, K. A. (1989). The finite-amplitude nature of tropical cyclogenesis. *J. Atmos. Sci.*, *46*(22), 3431–3456.

Emanuel, K. A. (2007, 11). Environmental factors affecting tropical cyclone power dissipation. *J. Climate*, *20*(22), 5497-5509.

Emanuel, K. A. (2008). The hurricane-climate connection. *Bull. Amer. Meteor. Soc.*, *89*, ES10-ES20.

Emanuel, K. A. (2013). Downscaling CMIP5 climate models shows increased tropical cyclone activity over the 21st century. *Proc Natl Acad Sci USA*, *110*, 12219–12224.

Emanuel, K. A. (2021). Atlantic tropical cyclones downscaled from climate reanalyses show increasing activity over past 150 years. *Nat. Commun.s*, *12*(1), 7027.

Emanuel, K. A. (2022). Tropical cyclone seeds, transition probabilities, and genesis. *J. Climate*, *35*(11), 3557–3566.

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. *Geosci. Model Dev.*, *9*, 1937–1958.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., … Zhang, M. (2011). The community climate system model version 4. *J. Climate*, *24*, 4973–4991.

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*, 474–479.

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., … Stössel, A. (2019). Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP). *Geosci. Model Dev.*, *12*(7), 3241–3281.

Haarsma, R. J., Acosta, M., Bakhshi, R., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., … Wyser, K. (2020). HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR – description, model computational performance and basic validation. *Geosci. Model Dev.*, *13*(8), 3507–3527.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., … von Storch, J.-S. (2016). High resolution model intercomparison project

(highresmip v1.0) for cmip6. *Geosci. Model Dev.*, *9*, 4185–4208.
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. *Nat. Commun.*, *11*, 3319.
Hegerl, G., & Zwiers, F. (2011). Use of models in detection and attribution of climate change. *WIREs Climate Change*, *2*(4), 570–591.
Hodges, K., Cobb, A., & Vidale, P. L. (2017). How well are tropical cyclones represented in reanalysis datasets? *J. Climate*, *30*(14), 5243–5264.
Jones, C. D., Hughes, J. K., Bellouin, N., Hardiman, S. C., Jones, G. S., Knight, J., ... Zerroukat, M. (2011). The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geosci. Model Dev.*, *4*, 543–570.
Klotzbach, P. J., & Gray, W. M. (2008). Multidecadal variability in North Atlantic tropical cyclone activity. *J. Climate*, *21*(15), 3929 - 3935.
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). *Bull. Amer. Meteor. Soc.*, *91*, 363–376.
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.), (p. 652-663). U.S. Global Change Research Program.
Knutson, T. R., Camargo, S., Chan, J., Emanuel, K., Ho, C.-H., Kossin, J., ... Wu, L. (2020a). Tropical cyclones and climate change assessment: Part I. detection and attribution. *Bull. Amer. Meteor. Soc.*, *100*, 1987–2007.
Knutson, T. R., Camargo, S. J., Chan, J. C. L., Emanuel, K., Ho, C.-H., Kossin, J., ... Wu, L. (2020b). Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bull. Amer. Meteor. Soc.*, *101*(3), E303 - E322.
Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K. A., Holland, G., Landsea, C., ... Sugi, M. (2010). Tropical cyclones and climate change. *Natural Geosci.*, *3*, 157-163.
Kossin, J. P. (2018). A global slowdown of tropical-cyclone translation speed. *Nature*, *558*, 104–107.
Kossin, J. P., Olander, T. L., & Knapp, K. R. (2013, August). Trend analysis with a new global record of tropical cyclone intensity. *J. Climate*, *26*, 9960–9976.
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. *J. Climate*, *33*, 4815-4834.
Lee, C.-Y., Sobel, A. H., Camargo, S. J., Tippett, M. K., & Yang, Q. (2022). New York State hurricane hazard: History and future projections. *J Appl Meteorol Climatol.*.
Lee, C.-Y., Tippett, M. K., Camargo, S. J., & Sobel, A. H. (2015). Probabilistic multiple linear regression modeling for tropical cyclone intensity. *Mon. Wea. Rev.*, *143*, 933-954.
Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2016). Autoregressive modeling for tropical cyclone intensity climatology. *J. Climate*, *29*, 7815-7830.
Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2018). An environmentally forced tropical cyclone hazard model. *J. Adv. Model. Earth Syst.*, *10*, 223-241.
Lloyd, E. A., & Oreskes, N. (2018, 2022/05/20). Climate change attribution: When is it appropriate to accept new methods? *Earth's Future*, *6*(3), 311–325.
Mann, M. E., & Emanuel, K. A. (2006, 2022/05/20). Atlantic hurricane trends linked to climate change. *Eos, Transactions American Geophysical Union*, *87*(24), 233–241.
Meiler, S., Vogt, T., Bloemendaal, N., Ciullo, A., Lee, C.-Y., Camargo, S. J., ... Bresch, D. N. (2022). Intercomparison of regional loss estimates from global

synthetic tropical cyclone models. *Nature Communications*, *13*(1), 6156.
Mizuta, R., Murata, A., Ishii, M., Shiogama, H., Hibino, K., Mori, N., ... Kimoto, M. (2017). Over 5,000 years of ensemble future climate simulations by 60-km global and 20-km regional atmospheric models. *Bull. Amer. Meteor. Soc.*, *98*(7), 1383–1398.
Moon, Y., Kim, D., Wing, A. A., Camargo, S. J., Zhao, M., Leung, L. R., ... Moon, J. (2022). An evaluation of tropical cyclone rainfall structures in the High-ResMIP simulations against satellite observations. *J. Climate*, 1–60.
Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., ... Tavoni, M. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168.
Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., ... Vidale, P. L. (2019). Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geosci. Model Dev.*, *12*(12), 4999–5028.
Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., ... Ullrich, P. (2020). Impact of model resolution on tropical cyclone simulation using the HighResMIP-PRIMAVERA multimodel ensemble. *J. Climate*, *33*, 2557–2583.
Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vannière, B., ... Wu, L. (2020). Projected future changes in tropical cyclones using the CMIP6 HighResMIP multimodel ensemble. *Geophys. Res. Lett.*.
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. *Mon. Wea. Rev.*, *142*, 3881-3899.
Sobel, A. H., Camargo, S. J., & Previdi, M. (2019). Aerosol versus greenhouse gas effects on tropical cyclone potential intensity and the hydrologic cycle. *J. Climate*, *32*(17), 5511–5527.
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. *Mon. Wea. Rev.*, *147*(7), 2355–2366.
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*, e2021EF00227.
Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., ... Miller, H. (2007). *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Tech. Rep.). Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.
Stone, D., Christidis, N., Folland, C., Perkins-Kirkpatrick, S., Perlwitz, J., Shiogama, H., ... Tadross, M. (2019). Experiment design of the international CLIVAR C20C+ detection and attribution project'. *Weather and Climate Extremes*, *24*, 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*, 70–74.
Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, *93*, 485–498.
Ting, M., Kushnir, Y., Seager, R., & Li, C. (2011, 2022/10/11). Robust features of Atlantic multi-decadal variability and its climate impacts. *Geophys. Res. Lett.*, *38*(17).
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. *J. Cli-*

mate, *21*, 2335-2357.

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. *Geosci. Model Dev.*, *10*, 1069–1090.

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. *Geosci. Model Dev.*, *2021*, 1–37.

Vecchi, G. A., Delworth, T. L., Murakami, H., Underwood, S. D., Wittenberg, A. T., Zeng, F., ... Yang, X. (2019). Tropical cyclone sensitivities to CO2 doubling: roles of atmospheric resolution, synoptic variability and background climate changes. *Climate Dynamics*, *53*(9), 5999–6033.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., ... Waldman, R. (2019, 2022/05/19). Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. *J. Adv. Model. Earth Syst.*, *11*(7), 2177–2213.

Wang, X., Wang, C., Zhang, L., & Wang, X. (2015). Multidecadal Variability of Tropical Cyclone Rapid Intensification in the Western North Pacific. *J. Climate*, *28*(9), 3806–3820.

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., ... Kimoto, M. (2010). Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *J. Climate*, *23*, 6312–6335.

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. *J. Climate*, *28*, 3905-3925.

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. *J. Adv. Model. Earth Syst.*, *6*, 980–997.

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).

Yan, X., Zhang, R., & Knutson, T. R. (2017). The role of Atlantic overturning circulation in the recent decline of Atlantic major hurricane frequency. *Nat. Commun.*, *8*(1), 1695.

Yoshida, K., Sugi, M., Mizuta, R., Murakami, H., & Ishii, M. (2017). Future changes in tropical cyclone activity in high-resolution large-ensemble simulations. *Geophys. Res. Lett.*, *44*(19), 9910–9917.

Zanchettin, D., Timmreck, C., Graf, H.-F., Rubino, A., Lorenz, S., Lohmann, K., ... Jungclaus, J. H. (2012). Bi–decadal variability excited in the coupled ocean–atmosphere system by strong tropical volcanic eruptions. *Climate Dynamics*, *39*, 419–444.

Zarzycki, C. M., & Ullrich, P. A. (2017). Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophys. Res. Lett.*, *44*(2), 1141–1149.

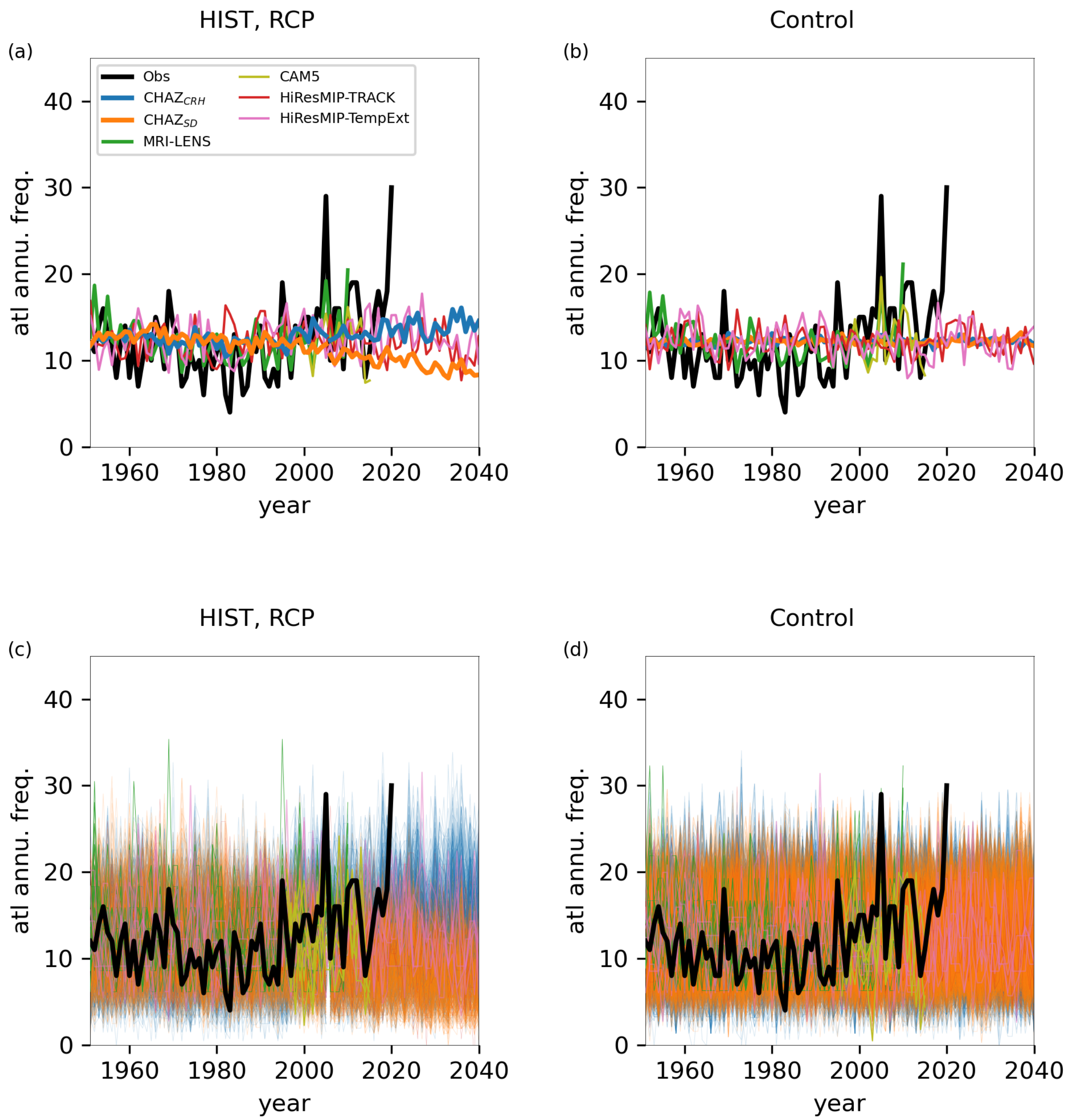

Figure 1. Annual frequency of Atlantic TCs exceeding 34 kt intensity threshold from 1951–2020 from best-track data (black), CMIP5 downscaling simulations using CHAZ$_{\mathrm{CRH}}$ (blue) and CHAZ$_{\mathrm{SD}}$ (pink), 25-km high-resolution CAM5 simulations (purple), 60km Japanese large-ensemble simulations (MRI-LENS, green), and HighResMIP simulations from (Roberts et al., 2020) and (Roberts et al., 2020). 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 CHAZ) 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.

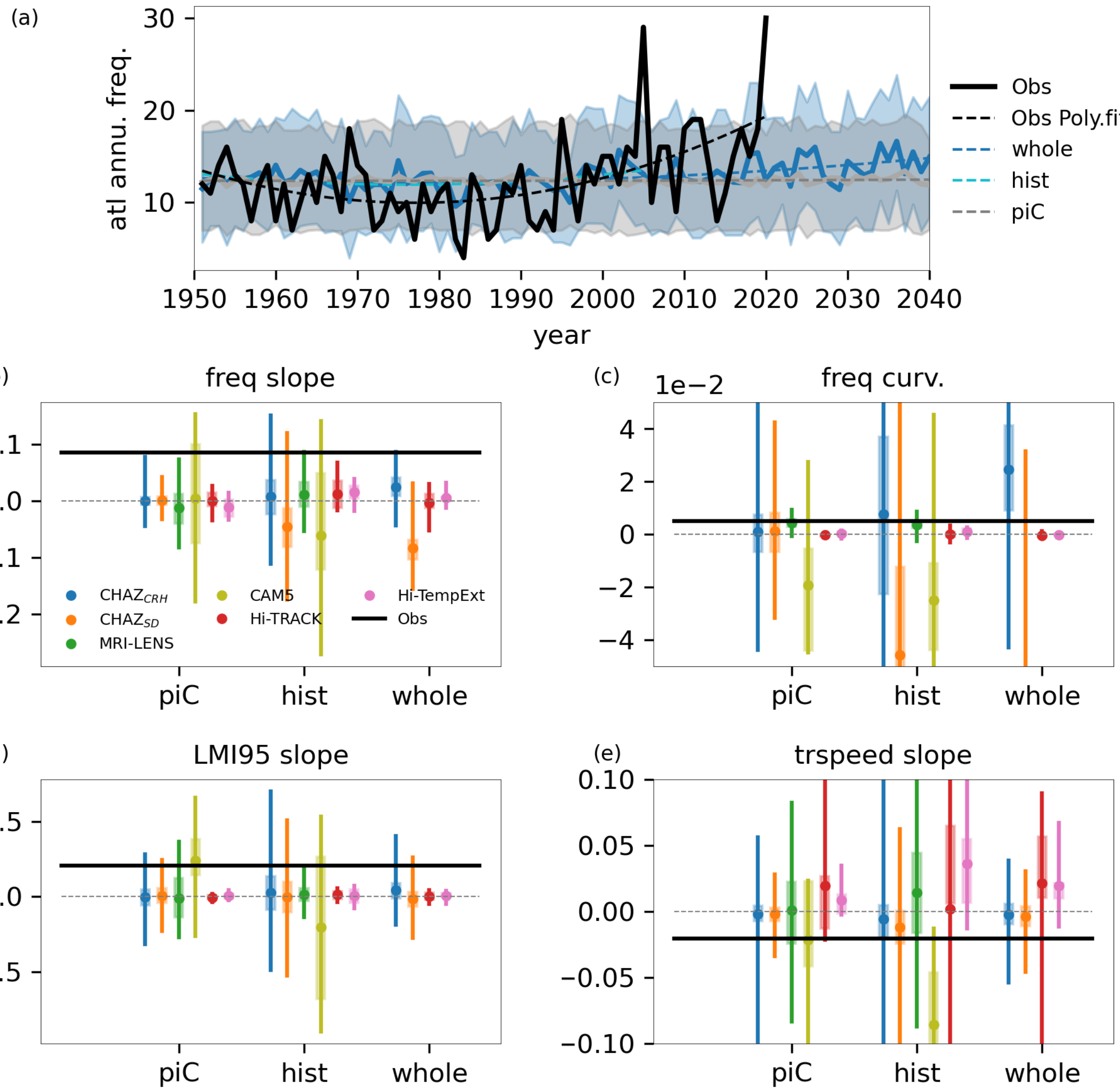

Figure 2. (a) Observed (black) and CHAZ$_{\mathrm{CRH}}$ simulated mean annual hurricane frequency. The CHAZ simulations are from present (1951-2005) to future climate (2006-2040) periods (blue), and from those using pre-industrial control climate forcing (gray). Dashed lines show the polynomial fit. '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 storms' annual frequency from all datasets. Datasets are indicated by color while the black line show the observed value. (c) Similar to (b) but for the quadratic terms. (d) and (e) are similar to (b) but for linear terms from he polynomial fit of LMI95 and storm forward motion 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}.

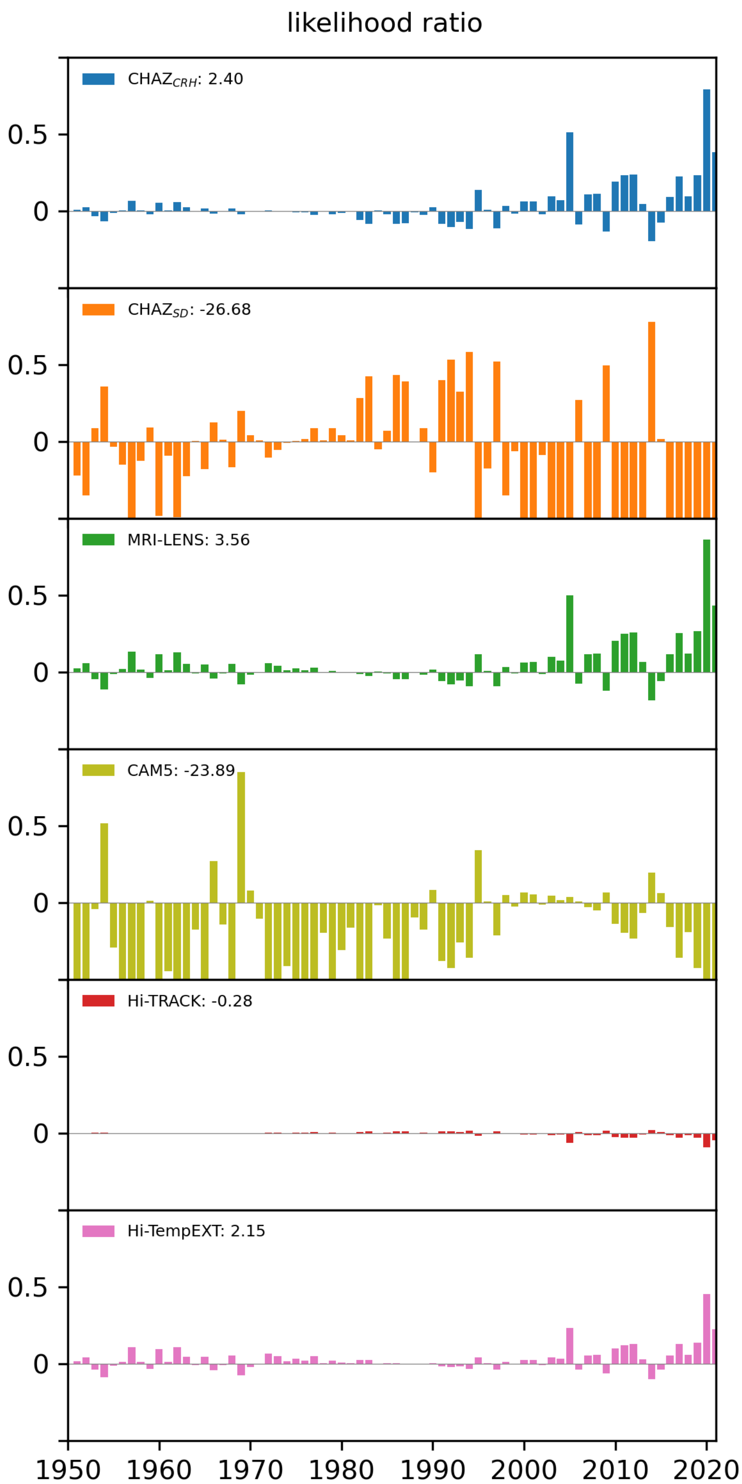

Figure 3. Annual log-likelihood ratio in which λ_t is derived from historical (and future for the CHAZ and HighResMIP runs) simulations and the annual likelihood that is estimated based on piC simulations.

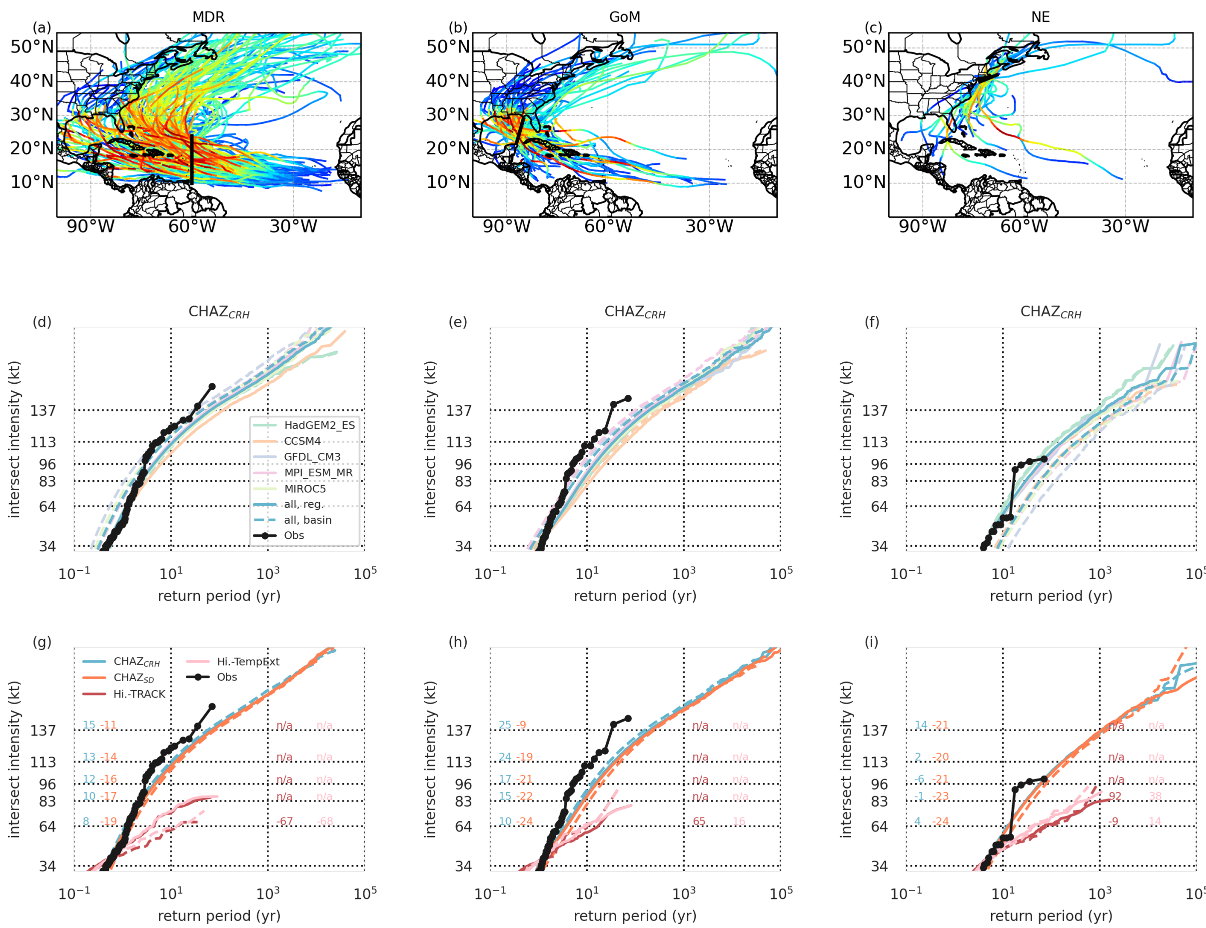

Figure 4. (a–c) Observed storm tracks from 1951–2020 at three line gates of interest. (d–f) Return period curves from 1951–2020 from best-track data (black lines), and $\mathrm{CHAZ}_{\mathrm{CRH}}$ historical simulations with basin-wide (dashed lines) and local (solid lines) basin corrections applied 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 datasets. 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. Datasets are indicated by colors. Black curves show the empirical return curve using observations from 1951–2020.

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], 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

Key Points:

- Changes in the Atlantic hurricane 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
- Based on our results, it is more likely that the risk of hurricanes is increasing than that it is decreasing, though not by a large margin

Corresponding author: Chia-Ying Lee, `cl3225@columbia.edu`

Abstract

This manuscript discusses the challenges in detecting and attributing recently observed trends in the Atlantic hurricanes and the epistemic uncertainty we face in assessing future hurricane risk. Data used here include 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 hurricane activity: the upward trend and multi-decadal oscillation of the annual frequency, the increase in storm wind intensity, and the downward trend in the forward speed. Some datasets suggest that these trends and oscillation are forced while others suggest that they can be explained by natural variability. Future 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 the future hurricane risk assessment. Here, we investigate the reduction of epistemic uncertainties on annual frequency through a statistical practice – likelihood analysis. We find that historical observations are more consistent with the simulations with increasing frequency but we are not able to rule out other possibilities. We argue that the most rational way to treat epistemic uncertainty is to consider all outcomes contained in the results. In the context of hurricane risk assessment, since the results contain possible outcomes in which hurricane 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 hurricanes. We looked at three aspects of these storms: the number of hurricanes 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 hurricane risk requires us to understand how the number of hurricanes 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 hurricanes 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 (Schreck et al., 2014; Emanuel, 2021; D. Chan et al., 2022). 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.

Accounting for climate change is likely to require a greater use of physics than is historically typical in catastrophe models (Toumi & Restell, 2014; Emanuel, 2008). 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., Klotzbach & Gray, 2008; J. C. Chan, 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 cyclones (TCs) 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 processes, 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), near future (up to 2040 or 2050), and pre-industrial control period. Simulations from pre-industrial control period contain no anthropogenic forcing signal 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 (tropical cyclone) 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 organically (Yoshida et al., 2017; Wehner et al., 2014; Roberts et al., 2020).

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 (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 aerosols over the Atlantic since 1980s has been argued to be a dominant cause of the increasing TC activity in late 20$^{\text{th}}$ century (Mann & Emanuel, 2006; Sobel, Camargo, & Previdi, 2019; Rousseau-Rizzi & Emanuel, 2020). On the other hand, several measures of Atlantic TC activity, including the major hurricane (TCs with LMI $\geq 93\,\text{kt}$) frequency (Goldenberg et al., 2001), are highly correlated to the the Atlantic Multi-decadal Oscillation (AMO) or Atlantic multidecadal variability (AMV), a low-frequency mode of variability identified by the average sea surface temperature anomalies in the North Atlantic basin, typically over 0-80$^{\text{o}}$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 (Clement et al., 2015; Bellomo et al., 2018). 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 the Atlantic TCs with warming (e.g., Knutson et al., 2010, and others), with a few exceptions (Emanuel, 2013; Bhatia et al., 2018; Vecchi et al., 2019).

The second objective is to compare historical simulations with observations to understand which modeling dataset 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 the global tropical cyclone genesis (i.e. Sobel et al., 2021). Lastly, we will assess hurricane risk over a set of selected 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 datasets

2.1.1 Observations

For reference, we use best-track data from National Hurricane Center obtained via International Best Track Archive for Climate Stewardship v04r00 IBTrACS (Knapp et al., 2010). We use 6-hourly storm positions (in longitude and latitude) and maximum wind speeds (kt) from 1951 to 2020. Storm forward speed is derived from the position data. We use only storms whose lifetime maximum intensity (LMI) reaches tropical storm (TS) strength, 34 kt. Hurricanes are referred to storms with LMI of at least 64 kt.

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 (tropical cyclone) 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. There is no feedback of 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 to 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 tropical cyclone 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) and Lee et al. (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. 2022). Meiler et al. (2022) found that losses estimated from CHAZ are comparable to those estimated using comparable academic tropical cyclone 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 (NCAR) Community Climate System Model 4 (CCSM4) (Gent et al., 2011), the Geophysical Fluid Dynamics Laboratory Climate Model version 3 (GFDL-CM3) (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 (MIROC5) (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 (Camargo et al., 2014). 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 thermodynamic inhibition of TC formation in a warming climate (Emanuel, 1989, 2022), but the gaps in our understanding of the relationship between climate and tropical cyclone 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 tropical cyclone 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 $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$.

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, i.e., tropical cyclone-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 was discussed in Yoshida et al. (2017). The second one is the 25-km High-Resolution Community Atmospheric Model version 5 simulations, CAM5 (Wehner et al., 2014, 2015). Next, we use storms tracked in the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016) High Resolution Model Intercomparison Project (HighResMIP) (Haarsma et al., 2016). Following Roberts et al. (2020) and Roberts et al. (2020), 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; Zarzycki & Ullrich, 2017; Ullrich et al., 2021), and both event sets are used here. 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 datasets. In the future period, $CHAZ_{CRH}$ and $CHAZ_{SD}$ contain storms from 2006-2040 under Representative Concentration Pathway 8.5 (rcp8.5) while HighResMIP storms are from 2015-2050 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) in ssp585 which considers a fossil-fueled development. 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 datasets as all years are equivalent. The MIR-LENS and CAM5 piC simulations are exceptions. In MRI-LENS and CAM5, the observed SST information is given in both historical and piC simulations as a lower boundary, but the long-term trend is removed in the piC simulations. In other words, MIR-LENS and CAM5 piC simulations still contain observed variation. 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 that only the linear trend is anthropogenic forced, not the variability, which, as we will discussed 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 4000 considering intensity ensemble), as is possible because the CHAZ intensity module has a stochastic component. Hi-TRACK has 7 members (5 global climate models and two of them have 2 ensemble members) and Hi-TempExt has 6 (4 global climate models and two of them have 2 ensemble members). MRI-LENS has 100 ensemble members while CAM5 has 5. The data properties are listed in Table 1.

2.3 Frequency adjustment

There are biases in 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-intensity biases, in general, as the grid spacings of these models are too coarse to capture the full range of observed hurricane strengths (e.g., Yoshida et al., 2017; Moon et al., 2022, and others). Here we address only the frequency biases. Specifically, we derive an adjustment by comparing the basin-wide annual TC frequency of models' historical simulations to that of the observations from the same period. The same adjustment will then be 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) and Lee et al. (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 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 dataset are from its respective historical simulation. Observations are calculated from their respective historical periods. 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 (in Section 5), we apply a multiplicative factor to ensure the annual frequency at storm with intensity greater than 40 kt in these data sets are consistent to observations, which is the same as the bias-correction approach used in (Lee et al., 2022).

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 rcp and hist 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 (Clement et al., 2015; Bellomo et al., 2018). 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 unites of the variable's unit per year (yr^{-1}) and per year square (yr^{-2}), respectively.

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). Figure 1a and b show the ensemble means of the time series of Atlantic hurricane frequency, i.e., the averaged total number of storms in the basin each year whose maximum sustained winds exceed 34 kt from each dataset. The small wiggles may be sampling variability. Figures 1c and d show the ensemble spread. By construction, the time-mean annual frequency for each dataset over its respective historical period will be identical to observations after the frequency adjustment (Eq. (1)). The original annual frequency of each dataset 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 $\mathrm{CHAZ_{SD}}$ (orange thick line) and $\mathrm{CHAZ_{CRH}}$ (blue thick line) results begin to diverge, with $\mathrm{CHAZ_{SD}}$ showing a decreasing TC frequency and $\mathrm{CHAZ_{CRH}}$ 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 HiResMIP datasets show no considerable trend in the historical period but a sharp dip after 2030. The ssp585 scenario in HiResMIP starts after 2015, though. Hi-TRACK's annual TC frequency climbs up by 2040. Roberts et al. (2020) 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 to the mean trend.

Figures 1b and 1d show analogous results for piC simulations. Note that the years in the x-axis are not real; 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 the Figure 1b, $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$ results do not diverge. There is no dip in the Hi-TRACK or Hi-TempExt. Clearly, the separation between $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$ and the dip in the two HighResMIP datasets in Figure 1a represent forced responses.

Next we conduct the trend analyses of the annual TC frequency in Figure 1 using second-order Legendre polynomials fits (Eq. (2)). As an example, Fig. 2a shows the analysis using the $\mathrm{CHAZ_{CRH}}$ simulations and the observations. The observed fit (dashed black line) has an upward trend of 0.085 storm year^{-2} and a positive curvature of 0.005 storm year^{-2} (shown as the black line in Figs. 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 (Clement et al., 2015; Bellomo et al., 2018). In Fig. 2a, the polynomial fits of $\mathrm{CHAZ_{CRH}}$ simulations from historical only (light blue dashed line) and from historical to future (dark blue dashed line) both show an small upward curve 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 dataset are also shown in Figures 2b-c. The observed linear trend are 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 lies within the 25-75 percentile ensemble ranges of piC simulations from $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$, and MRI-LENS. It is outside of the ensemble ranges from two HighResMIP datasets which have quadratic terms close to zero. The observed linear trend is at top 25 percentile (75-100 percentile) of the hist simulations of $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$, and is marginally included in the simulations of MRI-LENS; the observed quadratic term is within the 25-75 percentile range the $\mathrm{CHAZ_{CRH}}$ and MRI-LENS, and is at top 25 percentile in $\mathrm{CHAZ_{SD}}$. Only the fit linear trend derived from historical + future simulations of the $\mathrm{CHAZ_{CRH}}$ include the observed value. For the quadratic trend, the fit parameter derived from $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$ include the observed values but they are at top and bottom 25 percentile range, respectively. (We do not use any warming simulations from CAM5 and MRI-LENS.)

Generally speaking, the polynomial fit analysis suggests that, first, $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$ and MRI-LENS are better in capturing the observed trend and multi-decadal variability as their historical spread covers the observed values. However, CAM5 has only 10 years of data with 5 ensemble members and while Hi-TRACK and Hi-TempExt have only, respectively, 7 and 6 ensemble members. These three datasets may be under-sampled. Second, the observed linear trend is outside the spread of $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$ and MRI-LENS' piC simulations but within the spread of these models' hist simulations, indicating that anthropogenic forcing is necessary to capture the upward trend in the past decades. On the other hand, we can not rule out the possibility of the recent upward curvature trend is within the range of natural variability. Although the MRI-LENS' piC simulations is forced with the observed SST (with long-term trend removed) which results in the upward curvature term right on top of observed values in Figure 2c. Simulations from $\mathrm{CHAZ_{CRH}}$ suggest that that anthropogenic forcing helps to capture the upward curvature trend. Third, when considering the future period as well, the mean of $\mathrm{CHAZ_{CRH}}$ shows an upward trend, the mean of $\mathrm{CHAZ_{SD}}$ shows a downward trend, while the mean of the two HighResMIP simulations are close to zero. However, we have low confidence in the projections as they include zero. Thus, we can not say for sure that the positive linear and quadratic terms will continue into the future.

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

3.2 Intensity and storm motion

Figure 2d shows the fit parameters of Atlantic TC lifetime maximum intensity (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 linear term only. There is an upward trend in the observations, meaning that the extreme tail of observed intensity has increased with time, consistent with previous studies (e.g., Knutson et al., 2020a, and others). The positive linear trend is captured by the ensemble spreads of two CHAZ datasets and those of MRI-LENS and CAM5 at both piC and historical periods. It is outside of the ensemble spread of all simulations from from Hi-TRACK and Hi-TempExt. Thus, at least from $\mathrm{CHAZ_{CRH}}$, $\mathrm{CHAZ_{SD}}$, MRI-LENS, and CAM5, we can not rule out that the recent upward trend in the LMI95 is due to natural variability. When looking into the future, only the means of $\mathrm{CHAZ_{CRH}}$ is positive and the means of $\mathrm{CHAZ_{SD}}$, 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, indicating, again, low-confidence in the projected changes.

Figure 2e shows the analysis for translation speed. Consistent with (Kossin, 2018), the observations show a clear downward trend in the storm motion. This trend is within ensemble spread in all periods, including piC, for all models, except the simulations from Hi-TempExt. However, the mean and the 25-75 percentile ensemble spreads in these datasets move toward different directions from piC to hist to hist +future periods. The Hi-Track and MRL-LENS hist simulations show upward trends in the storm motion and this upward trends continues in to the future. The differences in mean and 25-75 percentile ensemble spreads from $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$ from these three period are small. The piC and hist simulations from CAM5 shows that anthropogenic forcing may lead to a strong downward trend in storm motion but again CAM5 simulations are shorter than do the data from 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 motion speed for TCs affecting Texas (Hassanzadeh et al., 2020) and an 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 trend in historical and historical + future vary from one dataset to another. This is especially true for the TC frequency projections between $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$, 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, multi-model ensemble mean is a common approach to obtain the consensus from multiple global climate models. However, such approach is only adequate when the ensemble spread represents variations that can be considered random, as might be the case with typical aleatoric uncertainties. The divergent scenarios in the frequency projections are a consequence of the epistemic uncertainty due to the lack of a satisfactory scientific understanding of tropical cyclone frequency (Sobel et al., 2021; Emanuel, 2022) 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 in 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 dataset by fitting the model annual TC frequency to a Poisson regression. We do so for all simulations with data throughout 2021 (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, i.e., 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 $\mathrm{CHAZ_{CRH}}$ simulations will generate 29 TCs as observed in 2005 is 0.08%, which is based on a Poisson distribution with $\lambda_{2005} = 15.7$. 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 larger than 0, the observations are more consistent with the simulations with anthropogenic forcing than with the piC simulations and vice versa.

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 $\mathrm{CHAZ_{CRH}}$, MRI-LENS, and Hi-TempExt. The annual likelihood ratios from these three datasets 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 datasets when compared to in piC (Figure 2a). As λ_t increases with time, the distributions from these three datasets 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, $\mathrm{CHAZ_{SD}}$ 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 $\mathrm{CHAZ_{SD}}$. 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 we consider the likelihood over the whole observational period, we calculate the average of the likelihood, i.e., the roots of $\prod_{2021}^{1950} L_t$ from all five datasets. Between $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$, observations are more consistent with $\mathrm{CHAZ_{CRH}}$, which has an averaged likelihood of 5.24%, than to $\mathrm{CHAZ_{SD}}$ which has the averaged likelihood of

3.46%. Among the five datasets, $CHAZ_{CRH}$ has highest likelihood, followed by Hi-TempExt (5.13%), MRI-LENS (5.1%), Hi-TRACK (5.04%), and CAM5 (3.6%).

The basin-wide frequency adjustment (Eq. (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 datasets 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. Without basin-wide TC frequency adjustment, Hi-TRACK has the greatest averaged likelihood, followed by CAM5, $CHAZ_{CRH}$, Hi-TempExt, $CHAZ_{SD}$, and MRI-LENS. MRI-LENS has the lowest likelihood because of its low bias and zero storms in some years.

5 Climate change and regional hurricane risk at three line gates

Now we compute regional hurricane risk, from hazard perspective only, represented by return periods of storms of given wind intensities passing through pre-defined gates, derived using simulations from historical and future periods. We use simulations from $CHAZ_{CRH}$, $CHAZ_{SD}$, Hi-TRACK, and Hi-TempExt. The three line gates used here (black lines in Figures 4a–c) are the main development region (MDR) gate which can be thought of as delineating Atlantic TC hazard in a general sense – how many storms form, and at what intensity and move from the MDR toward the US and Caribbean Islands; the GoM gate which records TC activity for those 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 (Eq. (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 adjustment shifts the simulated return period curves (solid line, local adjustment) by matching the return periods at 40 kt to the values derived from observations (see Section 2.3 for details). 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–i show the return period curves derived from the four datasets 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 (Roberts et al., 2020; Moon et al., 2022). 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, i.e., the differences between the dashed and solid lines, are small for the two CHAZ datasets in Figures 4g–i. 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-i. As expected, there is an overall increase in the storm frequency at all thresholds from historical to future periods for $\mathrm{CHAZ_{CRH}}$ while there is an overall decrease for $\mathrm{CHAZ_{SD}}$, consistent with the results in Figures 1 and 2a. The percentage changes are larger at higher intensity thresholds in the $\mathrm{CHAZ_{CRH}}$ but they are of similar or smaller magnitude throughout the Saffir-Simpson categories in the $\mathrm{CHAZ_{SD}}$. This is probably due to the increase in storm intensity as climate warms in $\mathrm{CHAZ_{CRH}}$ and $\mathrm{CHAZ_{SD}}$.

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+ ($\leq 64\,\mathrm{kt}$) 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+ storms, respectively. Hi-TempExt shows a 68% decrease and 16% increase of Category 1+ storms 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+ storms. 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 hurricane 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 $\mathrm{CHAZ_{CRH}}$ simulations and decreases in $\mathrm{CHAZ_{SD}}$, a difference that hangs on a subtle modeling choice (saturation deficit vs. relative humidity as a predictor of genesis). Changes in the high-resolution global climate model simulations are smaller, but again their direction depends on which global climate 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 dataset 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 $\mathrm{CHAZ_{CRH}}$, followed by Hi-TempExt, MRI-LENS, Hi-TRACK. However, the likelihood values among these four datasets are close to each other, so it would not be justified to draw a definitive conclusion from this analysis as to which model is most correct.

The epistemic uncertainty in CHAZ's projections on annual TC frequency comes from our design of the CHAZ model, but the conclusion is that our results are 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 attitude towards risk and the consequences of being wrong in either direction. 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 the present context, since the results contain possible outcomes in which TC risk — as estimated from a single model or subset of the entire multi-model ensemble — is increasing, that in itself means we should regard TC risk as increasing, at the highest level of understanding in which all available information is considered, even though there are other possible outcomes in which it is decreasing.

Open Research Section

CHAZ is an open-sourced model (`https://github.com/cl3225/CHAZ`). IBTrACS data are available at (`https://www.ncdc.noaa.gov/ibtracs/`). Information for CMIP5 data can be found at `https://pcmdi.llnl.gov/mips/cmip5/` and HighResMIP tropical cyclone information can be found at (`http://catalogue.ceda.ac.uk/uuid/e82a62d926d7448696a2b60c1925f81`. Underlying data for this publications are at (`https://github.com/cl3225/Lee_etal_2023EarthsFuture`).

Acknowledgments

Lee, Sobel, Camargo and Tippett were supported by Swiss Re MGMT LTD (CU18-3118). Wehner was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE340AC02-05CH11231 under the Regional and Global Model Analysis (RGMA) program.

References

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. *J Appl Meteorol Climatol.*, *49*(3), 493–520.

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.

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. *J. Climate*, *31*, 8281–8303.

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. *Nat. Commun.*, *10*(1), 635.

Bister, M., & Emanuel, K. A. (1997). The genesis of hurricane Guillermo: TEXMEX analyses and a modeling study. *Mon. Wea. Rev.*, *125*(10), 2662–2682.

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.

Bretherton, C. S., Peters, M. E., & Back, L. E. (2004). Relationships between water vapor path and precipitation over the tropical oceans. *J. Climate*, *17*, 1517-1528.

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.

Table 1. Data Characteristics

data	global climate models	resolution	ens	period	annual frequency
$\text{CHAZ}_{CRH/SD}$	HadGEM2_ES	N/A	100	1951–2005;2006–2040; piC	8.8 /15.9
	CCSM4				11/16.1
	GFDL_CM3				16.5/19.1
	MPI_ESM_MR				29.3/39.4
	MIROC5				11.9/18.3
MRI-LENS	MRI–AGCM3.2H	60 km	100	1950–2010; piC	2.3
CAM5	CAM5	28km	5	1996–2005;piC	10.9
Hi-TRACK	CMCC-CM2-VHR4 (r1i1p1f1)	25 km	7	1951-2014;2015-2040;piC	5.0
	CNRM-CM6-1-HR(r1i1p1f2)	50 km			21.0
	EC-Earth3P-HR (r1i1p2f1)	50 km			6.8
	EC-Earth3P-HR (r2i1p2f1)	50 km			6.5
	HadGEM3-GC31-HH (r1i1p1f1)	50 km			21.5
	HadGEM3-GC31-HM (r1i1p1f1)	50 km			19.7
	MPI-ESM1-2-XR (r1i1p1f1)	50 km			4.5
Hi-TempExt	CNRM-CM6-1-HR(r1i1p1f2)	50 km	6	1951-2014;2015-2040;piC	13.4
	EC-Earth3P-HR (r1i1p2f1)	50 km			2
	EC-Earth3P-HR (r2i1p2f1)	50 km			2
	HadGEM3-GC31-HH (r1i1p1f1)	50 km			13.3
	HadGEM3-GC31-HM (r1i1p1f1)	50 km			12.4
	MPI-ESM1-2-XR (r1i1p1f1)	50 km			0.63

Camargo, S. J., Giulivi, C., Sobel, A. H., Wing, A. A., Kim, D., Moon, Y., … Zhao, M. (2020). Characteristics of model tropical cyclone climatology and the large-scale environment. *J. Climate*, *33*, 4463–4487.

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. *J. Climate*, *27*, 9171-9196.

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).

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.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., … Navarra, A. (2019). Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model. *J. Adv. Model. Earth Syst.*, *11*(1), 185–209.

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*, 320–324.

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., … Zeng, F. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL Global Coupled Model CM3. *J. Climate*, *24*, 3484–3519.

Emanuel, K. A. (1989). The finite-amplitude nature of tropical cyclogenesis. *J. Atmos. Sci.*, *46*(22), 3431–3456.

Emanuel, K. A. (2007, 11). Environmental factors affecting tropical cyclone power dissipation. *J. Climate*, *20*(22), 5497-5509.

Emanuel, K. A. (2008). The hurricane-climate connection. *Bull. Amer. Meteor. Soc.*, *89*, ES10-ES20.

Emanuel, K. A. (2013). Downscaling CMIP5 climate models shows increased tropical cyclone activity over the 21st century. *Proc Natl Acad Sci USA*, *110*, 12219–12224.

Emanuel, K. A. (2021). Atlantic tropical cyclones downscaled from climate reanalyses show increasing activity over past 150 years. *Nat. Commun.s*, *12*(1), 7027.

Emanuel, K. A. (2022). Tropical cyclone seeds, transition probabilities, and genesis. *J. Climate*, *35*(11), 3557–3566.

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. *Geosci. Model Dev.*, *9*, 1937–1958.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., … Zhang, M. (2011). The community climate system model version 4. *J. Climate*, *24*, 4973–4991.

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*, 474–479.

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., … Stössel, A. (2019). Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP). *Geosci. Model Dev.*, *12*(7), 3241–3281.

Haarsma, R. J., Acosta, M., Bakhshi, R., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., … Wyser, K. (2020). HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR – description, model computational performance and basic validation. *Geosci. Model Dev.*, *13*(8), 3507–3527.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., … von Storch, J.-S. (2016). High resolution model intercomparison project

(highresmip v1.0) for cmip6. *Geosci. Model Dev.*, *9*, 4185–4208.
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. *Nat. Commun.*, *11*, 3319.
Hegerl, G., & Zwiers, F. (2011). Use of models in detection and attribution of climate change. *WIREs Climate Change*, *2*(4), 570–591.
Hodges, K., Cobb, A., & Vidale, P. L. (2017). How well are tropical cyclones represented in reanalysis datasets? *J. Climate*, *30*(14), 5243–5264.
Jones, C. D., Hughes, J. K., Bellouin, N., Hardiman, S. C., Jones, G. S., Knight, J., … Zerroukat, M. (2011). The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geosci. Model Dev.*, *4*, 543–570.
Klotzbach, P. J., & Gray, W. M. (2008). Multidecadal variability in North Atlantic tropical cyclone activity. *J. Climate*, *21*(15), 3929 - 3935.
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). *Bull. Amer. Meteor. Soc.*, *91*, 363–376.
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.), (p. 652-663). U.S. Global Change Research Program.
Knutson, T. R., Camargo, S., Chan, J., Emanuel, K., Ho, C.-H., Kossin, J., … Wu, L. (2020a). Tropical cyclones and climate change assessment: Part I. detection and attribution. *Bull. Amer. Meteor. Soc.*, *100*, 1987–2007.
Knutson, T. R., Camargo, S. J., Chan, J. C. L., Emanuel, K., Ho, C.-H., Kossin, J., … Wu, L. (2020b). Tropical cyclones and climate change assessment: Part II: Projected response to anthropogenic warming. *Bull. Amer. Meteor. Soc.*, *101*(3), E303 - E322.
Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K. A., Holland, G., Landsea, C., … Sugi, M. (2010). Tropical cyclones and climate change. *Natural Geosci.*, *3*, 157-163.
Kossin, J. P. (2018). A global slowdown of tropical-cyclone translation speed. *Nature*, *558*, 104–107.
Kossin, J. P., Olander, T. L., & Knapp, K. R. (2013, August). Trend analysis with a new global record of tropical cyclone intensity. *J. Climate*, *26*, 9960–9976.
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. *J. Climate*, *33*, 4815-4834.
Lee, C.-Y., Sobel, A. H., Camargo, S. J., Tippett, M. K., & Yang, Q. (2022). New York State hurricane hazard: History and future projections. *J Appl Meteorol Climatol.*.
Lee, C.-Y., Tippett, M. K., Camargo, S. J., & Sobel, A. H. (2015). Probabilistic multiple linear regression modeling for tropical cyclone intensity. *Mon. Wea. Rev.*, *143*, 933-954.
Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2016). Autoregressive modeling for tropical cyclone intensity climatology. *J. Climate*, *29*, 7815-7830.
Lee, C.-Y., Tippett, M. K., Sobel, A. H., & Camargo, S. J. (2018). An environmentally forced tropical cyclone hazard model. *J. Adv. Model. Earth Syst.*, *10*, 223-241.
Lloyd, E. A., & Oreskes, N. (2018, 2022/05/20). Climate change attribution: When is it appropriate to accept new methods? *Earth's Future*, *6*(3), 311–325.
Mann, M. E., & Emanuel, K. A. (2006, 2022/05/20). Atlantic hurricane trends linked to climate change. *Eos, Transactions American Geophysical Union*, *87*(24), 233–241.
Meiler, S., Vogt, T., Bloemendaal, N., Ciullo, A., Lee, C.-Y., Camargo, S. J., … Bresch, D. N. (2022). Intercomparison of regional loss estimates from global

synthetic tropical cyclone models. *Nature Communications*, *13*(1), 6156.
Mizuta, R., Murata, A., Ishii, M., Shiogama, H., Hibino, K., Mori, N., ... Kimoto, M. (2017). Over 5,000 years of ensemble future climate simulations by 60-km global and 20-km regional atmospheric models. *Bull. Amer. Meteor. Soc.*, *98*(7), 1383–1398.
Moon, Y., Kim, D., Wing, A. A., Camargo, S. J., Zhao, M., Leung, L. R., ... Moon, J. (2022). An evaluation of tropical cyclone rainfall structures in the HighResMIP simulations against satellite observations. *J. Climate*, 1–60.
Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., ... Tavoni, M. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168.
Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., ... Vidale, P. L. (2019). Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geosci. Model Dev.*, *12*(12), 4999–5028.
Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vanniere, B., ... Ullrich, P. (2020). Impact of model resolution on tropical cyclone simulation using the HighResMIP-PRIMAVERA multimodel ensemble. *J. Climate*, *33*, 2557–2583.
Roberts, M. J., Camp, J., Seddon, J., Vidale, P. L., Hodges, K., Vannière, B., ... Wu, L. (2020). Projected future changes in tropical cyclones using the CMIP6 HighResMIP multimodel ensemble. *Geophys. Res. Lett.*.
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. *Mon. Wea. Rev.*, *142*, 3881-3899.
Sobel, A. H., Camargo, S. J., & Previdi, M. (2019). Aerosol versus greenhouse gas effects on tropical cyclone potential intensity and the hydrologic cycle. *J. Climate*, *32*(17), 5511–5527.
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. *Mon. Wea. Rev.*, *147*(7), 2355–2366.
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*, e2021EF00227.
Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., ... Miller, H. (2007). *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Tech. Rep.). Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.
Stone, D., Christidis, N., Folland, C., Perkins-Kirkpatrick, S., Perlwitz, J., Shiogama, H., ... Tadross, M. (2019). Experiment design of the international CLIVAR C20C+ detection and attribution project'. *Weather and Climate Extremes*, *24*, 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*, 70–74.
Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, *93*, 485–498.
Ting, M., Kushnir, Y., Seager, R., & Li, C. (2011, 2022/10/11). Robust features of Atlantic multi-decadal variability and its climate impacts. *Geophys. Res. Lett.*, *38*(17).
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. *J. Cli-*

mate, *21*, 2335-2357.

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. *Geosci. Model Dev.*, *10*, 1069–1090.

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. *Geosci. Model Dev.*, *2021*, 1–37.

Vecchi, G. A., Delworth, T. L., Murakami, H., Underwood, S. D., Wittenberg, A. T., Zeng, F., ... Yang, X. (2019). Tropical cyclone sensitivities to CO2 doubling: roles of atmospheric resolution, synoptic variability and background climate changes. *Climate Dynamics*, *53*(9), 5999–6033.

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., ... Waldman, R. (2019, 2022/05/19). Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. *J. Adv. Model. Earth Syst.*, *11*(7), 2177–2213.

Wang, X., Wang, C., Zhang, L., & Wang, X. (2015). Multidecadal Variability of Tropical Cyclone Rapid Intensification in the Western North Pacific. *J. Climate*, *28*(9), 3806–3820.

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., ... Kimoto, M. (2010). Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *J. Climate*, *23*, 6312–6335.

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. *J. Climate*, *28*, 3905-3925.

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. *J. Adv. Model. Earth Syst.*, *6*, 980–997.

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).

Yan, X., Zhang, R., & Knutson, T. R. (2017). The role of Atlantic overturning circulation in the recent decline of Atlantic major hurricane frequency. *Nat. Commun.*, *8*(1), 1695.

Yoshida, K., Sugi, M., Mizuta, R., Murakami, H., & Ishii, M. (2017). Future changes in tropical cyclone activity in high-resolution large-ensemble simulations. *Geophys. Res. Lett.*, *44*(19), 9910–9917.

Zanchettin, D., Timmreck, C., Graf, H.-F., Rubino, A., Lorenz, S., Lohmann, K., ... Jungclaus, J. H. (2012). Bi–decadal variability excited in the coupled ocean–atmosphere system by strong tropical volcanic eruptions. *Climate Dynamics*, *39*, 419–444.

Zarzycki, C. M., & Ullrich, P. A. (2017). Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophys. Res. Lett.*, *44*(2), 1141–1149.

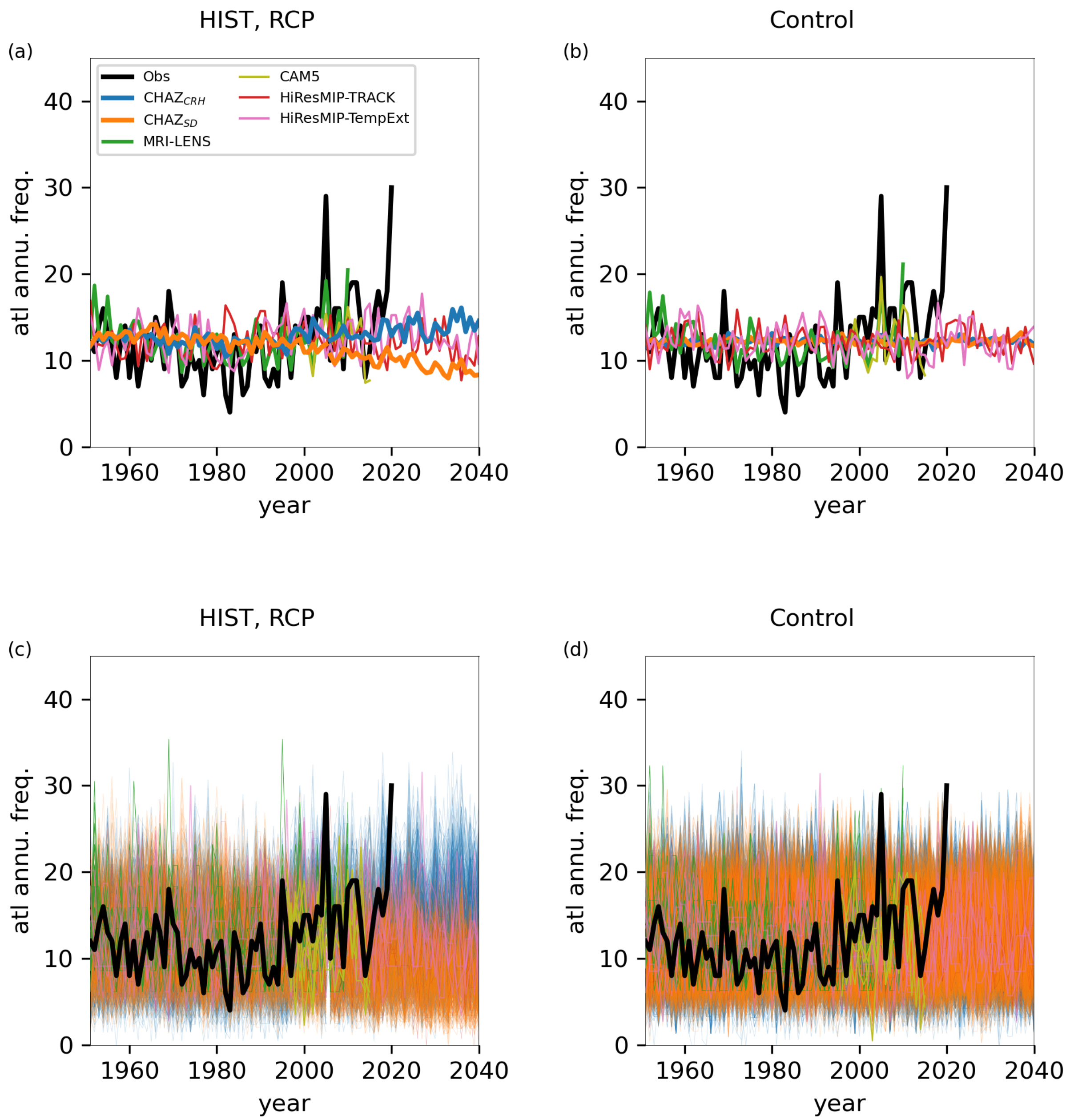

Figure 1. Annual frequency of Atlantic TCs exceeding 34 kt intensity threshold from 1951–2020 from best-track data (black), CMIP5 downscaling simulations using CHAZ$_{\mathrm{CRH}}$ (blue) and CHAZ$_{\mathrm{SD}}$ (pink), 25-km high-resolution CAM5 simulations (purple), 60km Japanese large-ensemble simulations (MRI-LENS, green), and HighResMIP simulations from (Roberts et al., 2020) and (Roberts et al., 2020). 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 CHAZ) 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.

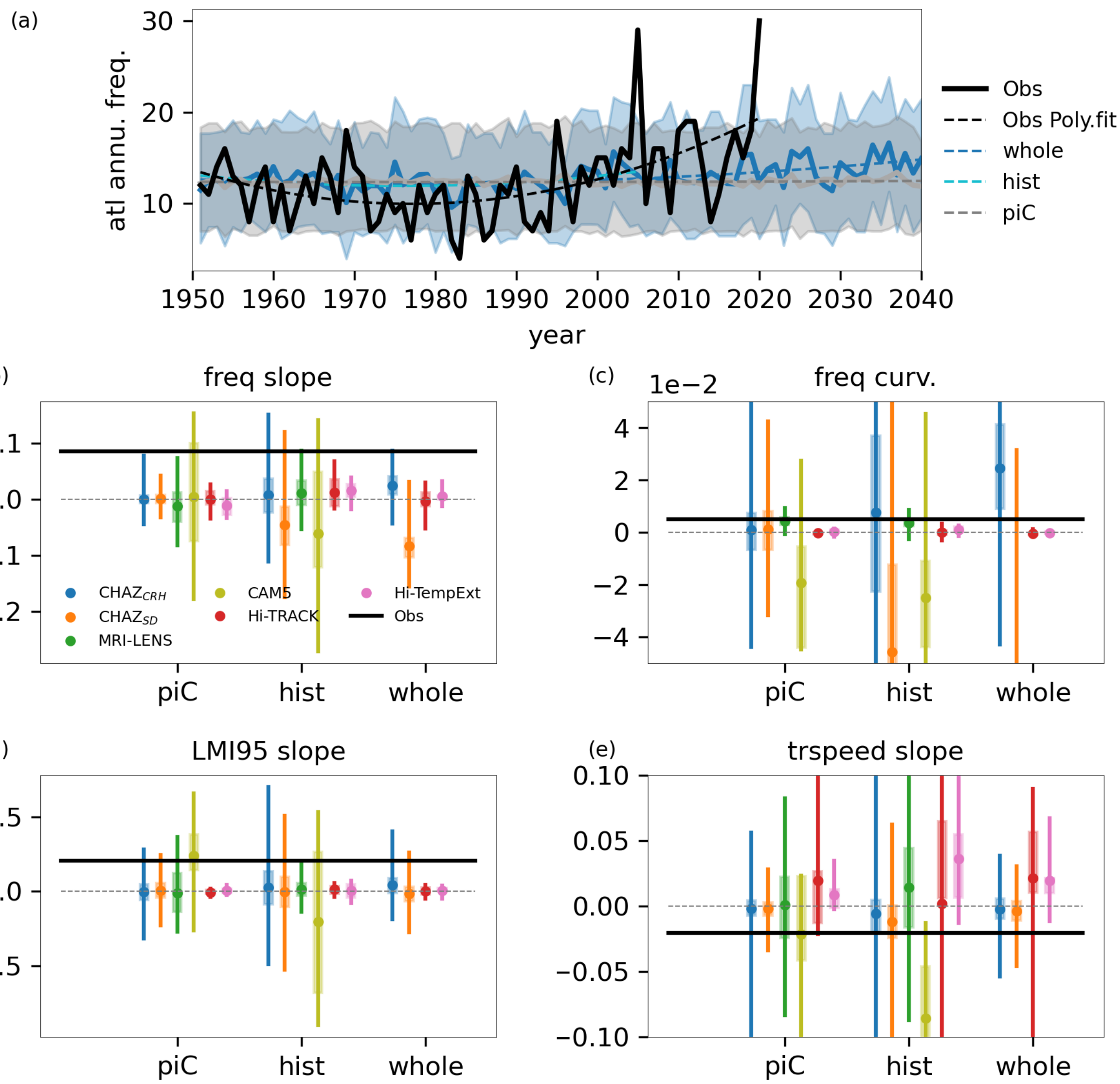

Figure 2. (a) Observed (black) and $\mathrm{CHAZ_{CRH}}$ simulated mean annual hurricane frequency. The CHAZ simulations are from present (1951-2005) to future climate (2006-2040) periods (blue), and from those using pre-industrial control climate forcing (gray). Dashed lines show the polynomial fit. '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 storms' annual frequency from all datasets. Datasets are indicated by color while the black line show the observed value. (c) Similar to (b) but for the quadratic terms. (d) and (e) are similar to (b) but for linear terms from he polynomial fit of LMI95 and storm forward motion 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}.

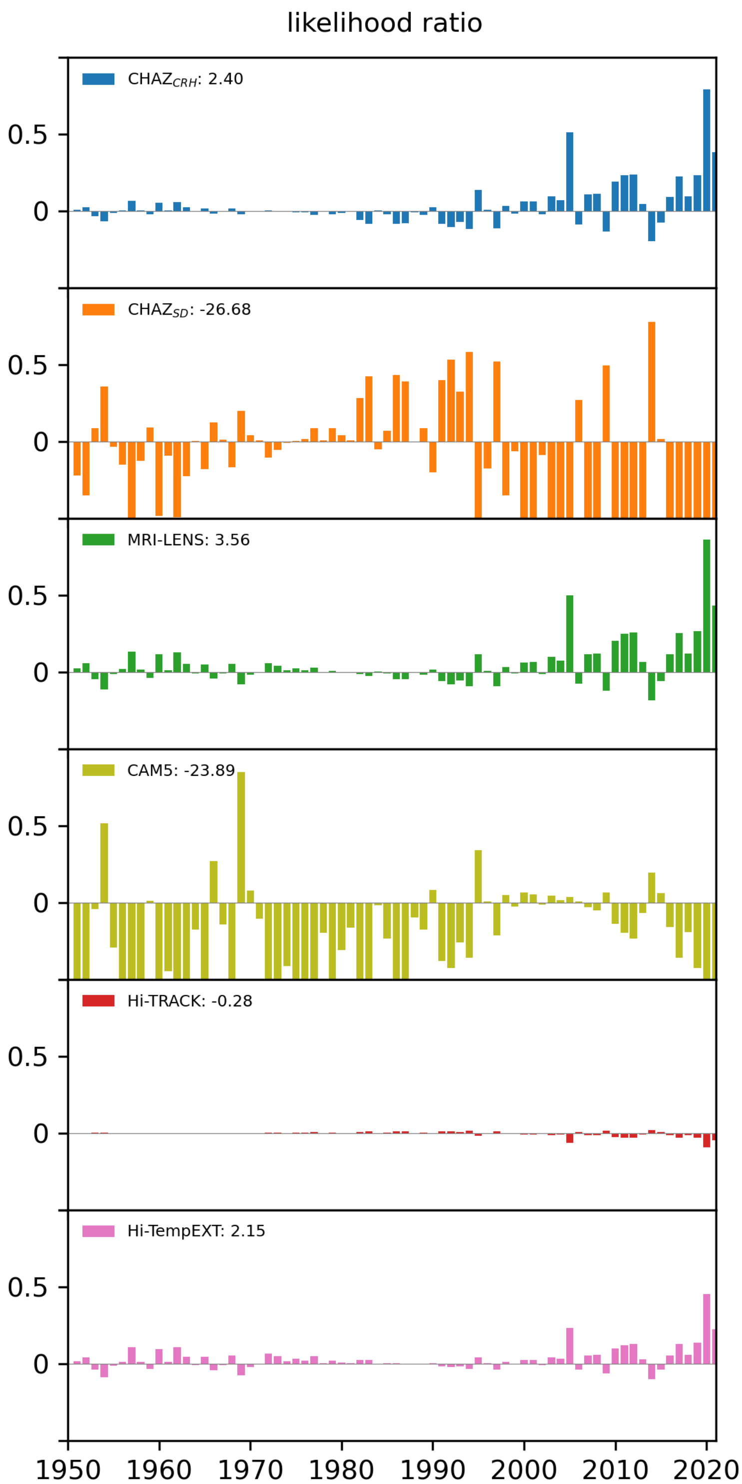

Figure 3. Annual log-likelihood ratio in which λ_t is derived from historical (and future for the CHAZ and HighResMIP runs) simulations and the annual likelihood that is estimated based on piC simulations.

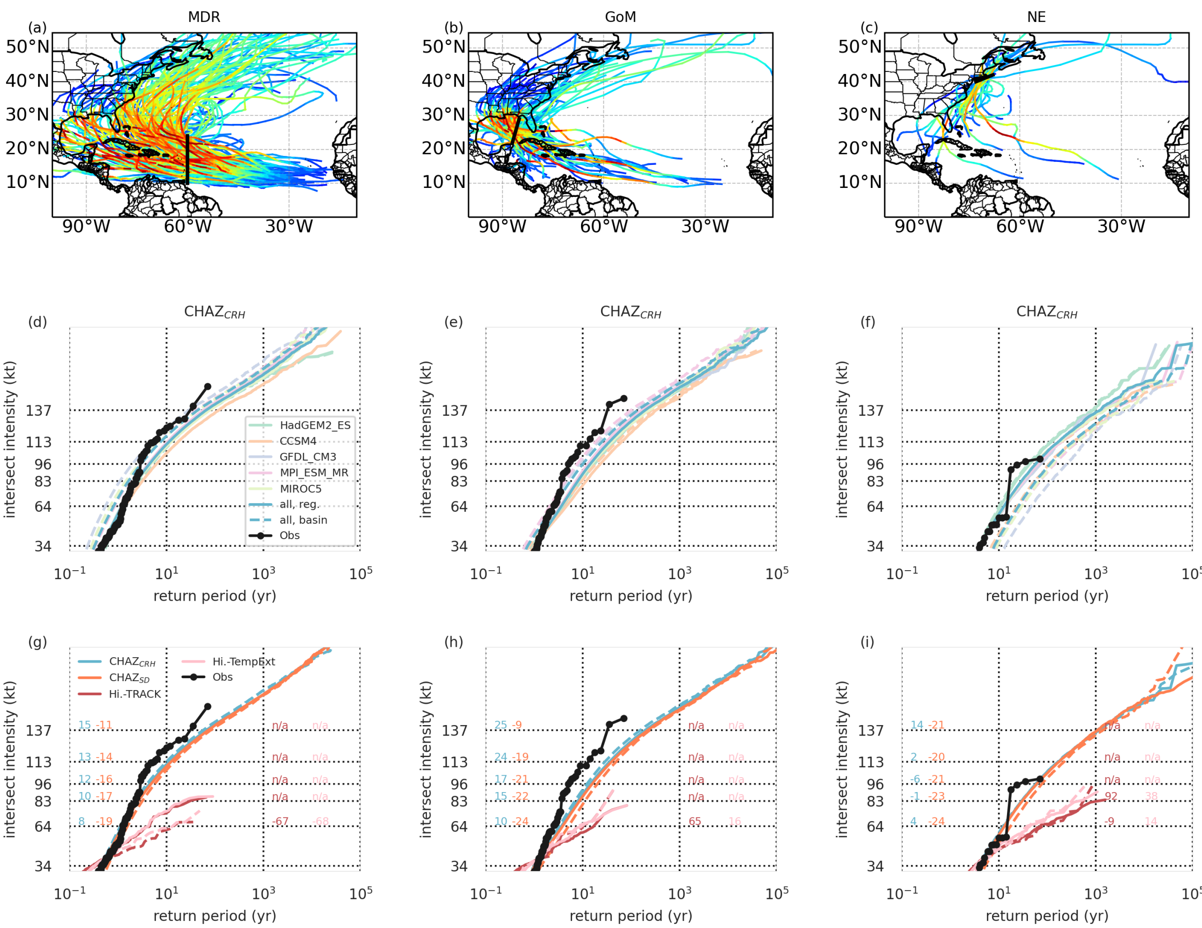

Figure 4. (a–c) Observed storm tracks from 1951–2020 at three line gates of interest. (d–f) Return period curves from 1951–2020 from best-track data (black lines), and $\mathrm{CHAZ_{CRH}}$ historical simulations with basin-wide (dashed lines) and local (solid lines) basin corrections applied at the 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 datasets. 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. Datasets are indicated by colors. Black curves show the empirical return curve using observations from 1951–2020.