Tropical cyclogenesis bias over the central North Pacific in CMIP6 simulations

Article

Accepted Version

Tropical Cyclogenesis Bias over the Central North Pacific in CMIP6 Simulations

Yi Peng,[a] Yi-Peng Guo,*[a] Jiuwei Zhao,[b] Zhe-Min Tan,[a] Xu Chen,[a] and Xiangbo Feng[c,d]

[a] *Key Laboratory of Mesoscale Severe Weather/Ministry of Education, and School of Atmospheric Sciences, Nanjing University, Nanjing, China*

[b] *Key Laboratory of Meteorological Disaster of Ministry of Education (KLME), Nanjing University of Information Science and Technology, Nanjing, China*

[c] *National Centre for Atmospheric Science and Department of Meteorology, University of Reading, Reading, UK*

[d] *Department of Physics, Imperial College London, UK*

Corresponding author: Yi-Peng Guo, guoyp@nju.edu.cn

Early Online Release: This preliminary version has been accepted for publication in *Journal of Climate*, may be fully cited, and has been assigned DOI 10.1175/JCLI-D-23-0224.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT

Current coupled climate models contain large biases in simulating tropical cyclogenesis, reducing the confidence in tropical cyclone (TC) projection. In this study, we investigated the influence of sea surface temperature (SST) biases on TC genesis in the Coupled Model Intercomparison Project Phase 6 simulations from 1979 to 2014. Positive TC genesis biases were found over the tropical Central North Pacific (CNP) in most of climate models, including the high-resolution models. Compared to coupled models, TC genesis density (TCGD) simulations over CNP in uncoupled models forced by observational SST improved obviously. A warm SST bias over the tropical CNP in the coupled models is the main cause of TC genesis biases. The SST bias-induced diabatic heating leads to an anomalous Gill-type atmospheric circulation response, which contributes to a series of favorable environmental conditions for TC formation over the CNP. Numerical experiments were also performed with HiRAM to demonstrate the influence of SST biases on the TCGD simulation, further confirmed our conclusion. The current results highlight the importance of improving TC simulation in state-of-the-art climate models by reducing SST simulation bias.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

1. Introduction

Tropical cyclone (TC) is one of the most disastrous weather systems that can lead to devastating winds, heavy rainfall, storm surge and huge economic losses upon landfall. With the ongoing threat of global warming, the prediction and long-term variability of TCs have become a hot topic of academic interest for a long time (McDonald et al. 2005; Li et al. 2010; Murakami et al. 2011a,b; Doi et al. 2013; Guo and Tan, 2018, 2022; Studholme et al. 2022). Benefited from the improvement of model performance, the state-of-the-art climate models have been widely used to study the long-term change of TC activities (Camargo and Wing 2015; Nakamura et al. 2017; Sharmila et al. 2020; Murakami and Wang 2022).

Tropical cyclogenesis is a basic aspect of TC activity in climate simulations, but this process contains large uncertainties (e.g., Tory et al. 2013; Shaevitz et al. 2014; Camargo et al. 2020; Sharmila et al. 2020). Sobel et al. (2021) provided a thorough and comprehensive review of research on the TC genesis in terms of observations, theories, model simulations, TC seeds and future research directions. In contrast to reanalysis data and TC best track data, models offer a longer timescale and more options for exploration, while series of idealized experiments can also be employed to help us uncovering the underlying physical mechanisms (Liu et al. 2012; LaRow et al. 2014; Zhao et al. 2020). However, TC biases arising from coarse resolution (e.g., Camargo et al. 2020; Roberts et al. 2020), inaccurate physical parameterization (e.g., Vitart et al. 2001; Kim et al. 2012), and sea surface temperature (SST) biases (e.g., Hsu et al. 2018; Chan et al. 2021) always bring large uncertainties to TC simulations. SST is an essential factor that impacts TC genesis simulations in climate models (e.g., Strazzo et al. 2013; Hsu et al. 2018; Dutheil et al. 2020; Zhang et al. 2021). For instance, the cold SST biases could be one of reasons for underestimated TCs in the North Atlantic (Kim et al. 2014; Sharmila et al. 2020), while inappropriate representation of monsoon characteristics was also found to be responsible for the errors in the annual cycle of simulated TC activity in the North Indian Ocean (Tory et al. 2020). A better simulation of TC activity in climate models could help to improve both TC seasonal prediction and future climate projection.

SST biases exist at large across the current climate models in different ocean basins, with the most significant warm bias on the east coast of the Pacific Ocean (Richter 2015), which is an important source of uncertainty in TC simulations under climate change (Vecchi et al. 2019). Dutheil et al. (2020) used an emergent-constraint method to correct the model SST

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

bias in the South Pacific, substantially reducing cyclogenesis bias. This indicates that the uncertainty in the projected SST patterns strongly affects the reliability of TC projections. Hsu et al. (2018) identified the key regions in which reducing SST biases could potentially improve TC representation in climate models. Meanwhile, efforts have also been made to correct the SST bias. Vecchi et al. (2014) attempted to use flux adjustment to correct the SST and the effects on TC simulation were remarkable. Chan et al. (2021) applied recently developed corrections for biases in historical SSTs, which lead to revisions in tropical to subtropical SST gradients and improved simulation of TC activity.

The Coupled Model Intercomparison Project (CMIP) offers a large number of models and experiments, which provide a good testbed to evaluate the TC simulation in current climate models. Previous evaluations of different models in Phase 3 or 5 of the CMIP (CMIP3/5) have revealed future changes in TC activity, but obvious uncertainties also exist (Bell et al. 2019; Camargo 2013; Kossin et al. 2016). The latest launched Coupled Model Intercomparison Project Phase 6 (CMIP6) includes most of the current state-of-the-art climate models on which future climate projections rely. Previous studies have revealed that the enhancement of model resolution can improve TC simulation (Nakamura et al. 2017; Camargo et al. 2020). Especially, the High Resolution Model Intercomparison Projection (HighResMIP; Haarsma et al. 2016) in CMIP6 performs well in simulating TC characteristics, such as intensity and rainfall due to their high spatial resolutions (Roberts et al. 2020; Zhang et al. 2021). However, obvious SST biases still exist in most of the CMIP6 models, as they did in the previous CMIP5 models (Han et al. 2021; Zhang et al. 2023), which may affect TC simulation.

Compared with atmospheric general circulation models (AGCMs) that use observed SSTs, coupled atmosphere-ocean general circulation models (AOGCMs) with SST biases can have some impacts on the simulation of TCs to some extent. For example, subsurface cold water plays a key role in reducing biases in the distribution of intense TCs in AGCMs (Ogata et al. 2015, 2016). Li and Sriver (2018) also confirmed this finding and further indicated that the key differences in storm number and distribution can be attributed to variations in the modeled large-scale climate mean state and variability that arises from the combined effect of intrinsic model biases and air-sea interactions. Morim et al. (2020) found that the surface wind bias is largely intrinsic to the atmospheric components of the models and inconsistencies between AGCM and AOGCM simulations are mainly driven by SST errors.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

Finally, model evaluation of the ability of GCMs or multi-model ensemble to simulate TCs is also an essential aspect, and some studies also mentioned the impact of SST on this evaluation (Bengtsson et al. 2007; Camargo et al. 2016; Balaguru et al. 2020; Wu et al. 2021). Sobel et al. (2021) emphasized the critical influence of atmospheric-ocean coupling and model resolution on the accurate simulation of TC genesis. Meanwhile, they also highlight that besides the large-scale environmental fields, TC seeds in climate models also have a great influence on TC frequency.

The North Pacific accounts for nearly 70% TCs over the entire Northern Hemisphere. It remains unclear how the current state-of-the-art CMIP6 models perform in simulating TC genesis over the North Pacific (Patricola et al., 2022), given that CMIP5 shows large TC genesis bias over this region (Camargo 2013; Tory et al. 2020). The HighResMIP models, which have relatively higher model resolutions, also provide new opportunities for TC research, enabling the identification of biases. Wang et al. (2014) found cold SST biases over the CNP in CMIP5 models, while Zhang et al. (2023) revealed that the SST bias in CMIP6 models underwent a warming change relative to CMIP5 models, including the CNP. The instability of SST simulations in climate models is likely to affect TC generation simulations over this region. Thus, whether the SST bias plays a role in inducing TC simulation bias also deserves further investigation.

The remainder of this paper is organized as follows: Section 2 describes the data, models and methods used in this study, Section 3 presents the results in different experiments from CMIP6, Section 4 gives the analysis of large-scale environmental conditions and the numerical experiment, and Section 5 discusses the results and formulates conclusions.

2. Data and methods

a. Observational data

The TC best track dataset was obtained from the Joint Typhoon Warning Center (JTWC) in the International Best Track Archive for Climate Stewardship version 4 (IBTrACS V4; Knapp et al. 2010). TCs with the maximum intensity exceeding the tropical storm intensity (surface sustained wind speeds $\geq$ 34 kt) were considered in this study. TC genesis was defined as the first record with intensity above 34 kt. The European Center for Medium Range Weather Forecasts (ECMWF) monthly reanalysis 5 (ERA5) with 0.25° horizontal

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

resolution (Hersbach et al. 2020) was used to analyze the large-scale environmental fields. The monthly mean SST data were derived from the Hadley Centre Sea Ice and Sea Surface Temperature (HadISST; Rayner et al. 2003) with 1° horizontal resolution.

b. CMIP6 simulations

The CMIP6 model output data used in this study were from four simulations, namely "historical", "AMIP", "hist-1950" and "highresSST-present". The historical simulation is coupled runs forced by externally imposed conditions like solar variability and volcanic aerosols based on historical observations over the period 1850–2014. The AMIP experiment employs the observational data based on the merged Hadley-OI sea surface temperature and sea ice concentration datasets since 1979 as boundary conditions to drive uncoupled atmospheric circulation models. The hist-1950 and highresSST-present simulations, on the other hand, belong to the HighResMIP experiments for 1950–2014 (Haarsma et al. 2016), which have higher horizontal resolutions (25–50km). "hist-1950" simulations are coupled runs, while the "highresSST-present" simulations are atmospheric models forced by SST from HadISST and sea-ice datasets. In this study, 16 "historical", 10 "AMIP", 10 "hist-1950", and 10 "highresSST-present" models were chosen. It should be emphasized that the same 10 models were chosen from both hist-1950 and highresSST-present simulations to investigate the effect of SST and air-sea interaction on model TC simulations (e.g., ECMWF-IFS-HR from both hist-1950 and highresSST-present simulations). The first realizations of these model simulations were analyzed and more details about these models can be found in Table 1. We used 6-hourly data for TC detection and monthly data for analyzing the environmental fields connected with the simulated TC activity. 10 coarse-resolution uncoupled atmospheric models from the AMIP experiment, which corresponds to the historical models, are only used in the Section 3c for comparison.

Experiments	Model name	Resolution (lon×lat)
historical	ACCESS-ESM1-5	192×144
	BCC-CSM2-MR	320×160
	CMCC-ESM2	288×192
	CNRM-CM6-1-HR	720×360
	EC-Earth3	512×256
	FGOALS-f3-L	288×180
	GFDL-ESM4	288×180
	GISS-E2-1-G	144×90

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

	HadGEM3-GC31-MM	432×325
	KIOST-ESM	192×96
	MIROC6	256×128
	MRI-ESM2-0	320×160
	NESM3	192×96
	NorESM2-MM	288×192
	SAM0-UNICON	288×192
	TaiESM1	288×192
AMIP	ACCESS-ESM1-5	192×144
	BCC-CSM2-MR	320×160
	CNRM-CM6-1-HR	720×360
	EC-Earth3	512×256
	HadGEM3-GC31-MM	432×325
	MIROC6	256×128
	MRI-ESM2-0	320×160
	NESM3	192×96
	SAM0-UNICON	288×192
	TaiESM1	288×192
hist-1950	CMCC-CM2-HR4	288×192
	CMCC-CM2-VHR4	1152×768
	CNRM-CM6-1-HR	720×360
	CNRM-CM6-1	256×128
	EC-Earth3P	512×256
	EC-Earth3P-HR	1024×512
	ECMWF-IFS-HR	720×361
	ECMWF-IFS-LR	360×181
	HadGEM3-GC31-HM	1024×768
	HadGEM3-GC31-MM	432×324
highresSST-present	CMCC-CM2-HR4	288×192
	CMCC-CM2-VHR4	1152×768
	CNRM-CM6-1-HR	720×360
	CNRM-CM6-1	256×128
	EC-Earth3P	512×256
	EC-Earth3P-HR	1024×512
	ECMWF-IFS-HR	720×361
	ECMWF-IFS-LR	360×181
	HadGEM3-GC31-HM	1024×768
	HadGEM3-GC31-MM	432×324

Table 1. Details of CMIP6 models used in this study, including experiments, names, institutions and model resolution.

c. TC detection algorithm

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

The TC detection algorithm used in this study is the TSTORMS method developed by the Geophysical Fluid Dynamics Laboratory (GFDL) (M. Zhao et al. 2009; J. Zhao et al. 2020; Burnett et al. 2021; Song et al. 2022). The algorithm requires sea level pressure, 850-hPa relative vorticity, near-surface wind, and warm core to detect TCs. First, the algorithm uses thresholds for the distance between the center of the minimum pressure of the vortex and the warm core, the magnitude of the relative vorticity and the intensity of the warm core to extract a single vortex for each time period (every six hours) in the model. Once the vortexes are extracted, they are tied together to form a trajectory if they reach the minimum wind speed and warm core intensity, and the distance between two reports should be less than the threshold. Finally, only the vortex that forms a section of the trajectory that reaches the lifetime maximum intensity threshold within a defined lifespan will be extracted as a TC.

We used the 850-hPa wind speed instead of the surface wind speed in the historical models for the coarse resolution, while we used the surface wind speed in the HighResMIP models and HiRAM (Camargo et al. 2013; Song et al. 2022). We averaged the 250-, 500-, and 850-hPa temperatures to identify the warm core, because all models have at least these three levels. Limited by the model resolution, the TCs in most models are relatively weak, and the genesis frequency is low compared to the observation. Therefore, the thresholds are relaxed when detecting TCs in historical simulations, as has been done in previous studies (Camargo et al. 2013; Roberts et al. 2020). The thresholds in historical models were chosen in order to get closer TC frequency comparing to the observation. But for the same model in hist-1950 and highresSST-present simulations, we used uniform thresholds (e.g., ECMWF-IFS-HR from both hist-1950 and highresSST-present simulations). In addition, for all the simulations, the lifespan of the vortexes was required to exceed two days (eight reports). The threshold for relative vorticity was larger than 3×10^{-5} s^{-1} and the distance between two reports cannot exceed 700km, more details about the thresholds can be found in Table 2. It needs to be emphasized that, our results weren't affected by different detection algorithms such as TRACK, which will be further discussed in Section 3d. The conclusion also isn't affected by altering the thresholds slightly, even if the criterion of TC genesis changes to the 17.2m/s (34kt) in all the models from the three experiments.

Model	Minimum wind speed (m/s)	Lifetime maximum wind speed (m/s)	Warm core intensity (℃)
ACCESS-ESM1-5	10	14	0.3
BCC-CSM2-MR	17	17	0.8
CMCC-ESM2	8	14	0.3
CNRM-CM6-1-HR	8	14	0.5

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

EC-Earth3	8	14	0.3
FGOALS-f3-L	17	17	1.0
GFDL-ESM4	17	17	1.0
GISS-E2-1-G	8	10	0.3
HadGEM3-GC31-MM	12	17	0.8
KIOST-ESM	15	17	0.5
MIROC6	8	14	0.3
MRI-ESM2-0	10	17	0.5
NESM3	8	14	0.3
NorESM2-MM	8	10	0.5
SAM0-UNICON	17	17	1.0
TaiESM1	8	14	0.3

Table 2. Details of thresholds used in TSTORMS to detect TCs in 16 CMIP6 historical models.

d. HiRAM model

The High Resolution Atmospheric Model (HiRAM; Zhao et al. 2009) developed by GFDL was used to perform two experiments with different SST forcing running from 1980 to 2009 for 30 years to verify the mechanism in our study. The HiRAM has a horizontal resolution of ~60 km and 32 vertical levels and uses a finite-volume core by a cubed-sphere grid topology as the dynamical core (Putman and Lin 2007). It can be used to explicitly extract TC-like vortices and well represent the TC characteristics and climatology. Numerous past studies have used HiRAM for TC activities (M. Zhao et al. 2009, 2010; Camargo et al. 2014; Knutson et al. 2015; J. Zhao et al. 2020; Chen et al. 2023)

e. Statistics and definitions

Tropical cyclone genesis density (TCGD) in every 2.5° grid box was calculated in both observation and models. TC genesis in the observed data was defined as intensity exceeding the tropical storm intensity, while in the models it was defined as the first report detected by the TSTORMS (reaching every threshold). The relative TCGD in this study was defined as the proportion of TCGD of each grid point over the North Pacific relative to the total TCGD of the whole North Pacific.

Due to the inherent variations in the frequency of TCs simulated by different models. A direct comparison of absolute TCGD may lead to misinterpretations and thus does not provide a comprehensive assessment of the actual scenario. Therefore, normalizing the

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

TCGD values with respect to the total TCGD of the North Pacific was performed to eliminate the biases introduced by inter-model discrepancies in the simulated TC frequency.

All the monthly data above were interpolated to 2.5° × 2.5° grid using bilinear interpolation in the analysis. The study period is 1979–2014 for both the observations and the CMIP6 simulations. The TC season was defined as June–November (JJASON), and a Student's *t*-test was applied to evaluate the statistical significance.

In comparing the results of coupled and uncoupled models, we employed a significance test based on sign similar to Chand et al. (2017). Statistical robustness is determined when there is consistent agreement among a significant proportion of models regarding the sign of the difference according to the binomial distribution.

3. TC genesis biases in CMIP6 models

Figure 1a shows the climatological distribution of JJASON TCGD in the observations over the North Pacific region during 1979–2014. We can clearly see that there are two main genesis regions for TCs, WNP and ENP, while the Central North Pacific region (170°E~150°W, 0°~30°N) in the red box displayed less TC formation, attributed to the excessive dry air and the strong VWS associated with the Tropical Upper Troposphere Trough (TUTT) (Wang and Wu 2018; Tory et al. 2020). However, in the result of multi-model ensemble mean (MME), calculated by all CMIP6 models, the TC genesis region is more dispersed and the TCGD over the ENP and WNP is significantly lower than observed, while more TCs are genesis over the CNP region (Fig. 1b).

The MMEs of historical, hist-1950 and highresSST-present models used in this study were also calculated, respectively. We found that positive TCGD biases mainly came from coupled historical and hist-1950 models (Figs. 1c,d). Comparing historical models with normal resolution and high-resolution hist-1950 models, we can see that the increase in resolution improves the positive TCGD bias of the model to some extent, but a considerable and significant positive bias still exists over the CNP region. Nevertheless, comparing hist-1950 and highresSST-present models from HighResMIP, such positive bias is significantly reduced (Figs. 1d,e). Previous studies have revealed that most of the deficiencies in TC simulations in the ENP and WNP stem from a lack of model resolution and narrow basin (Bengtsson et al. 2007; Davis 2018), while positive TCGD biases in the CNP region have rarely been noted.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

It is worth noting that the modeled TC activity may be affected the TC detection methods. To validate whether the overestimation of CNP TCGD relies on the TSTORMS method, we compared the extracted TCs in ERA5 reanalysis by using both TSTORMS and the Okubo-Weiss-Zeta (OWZ) method (Chand et al. 2017, 2022). It turns out that both methods keep good consistency in extracting climatological TC distribution from ERA5. The overestimation of the CNP TCGD in historical simulation is also reproduced by the OWZ method (not shown).

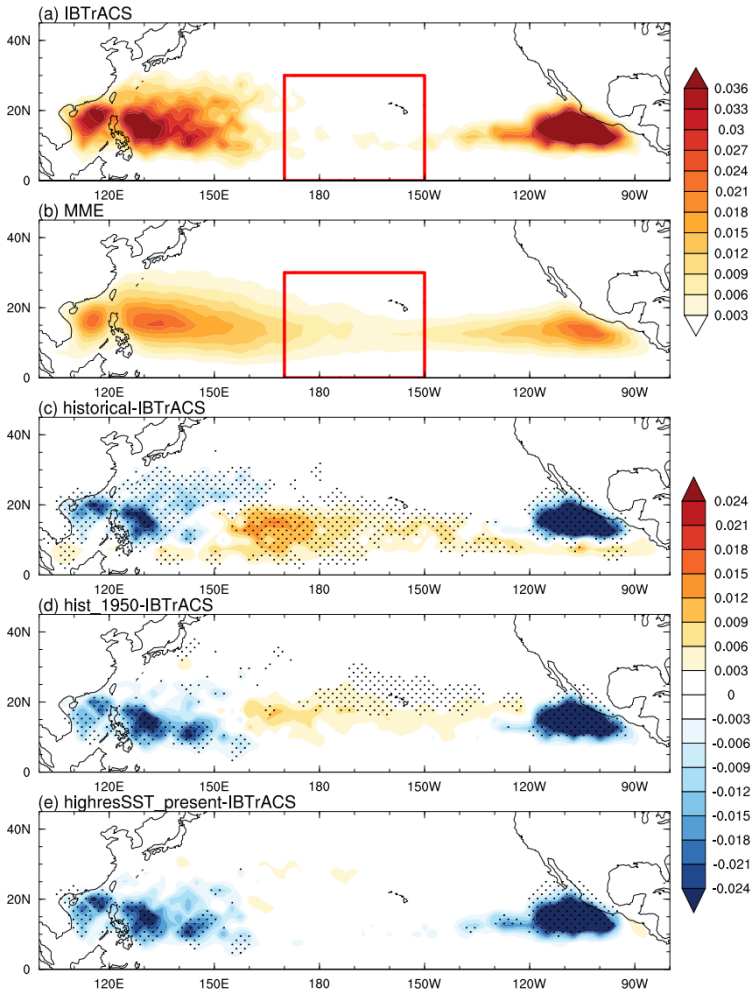

Fig. 1. Climatological distributions of the JJASON TCGD in (a) IBTrACS and (b) multi-model ensemble mean of all CMIP6 models. TCGD difference between IBTrACS and the multi-model ensemble mean of (c) historical, (d) hist-1950, (e) highresSST-present simulations during 1979–2014. The red boxes indicate the CNP region (170°E–150°W, 0°N–30°N). Only values exceeding the 95% confidence level are stippled, with at least 11 out of 16 models in (c) and 8 out of 10 models in (d–e) agreeing on the sign of difference.

a. TCGD in historical runs

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0224.1.

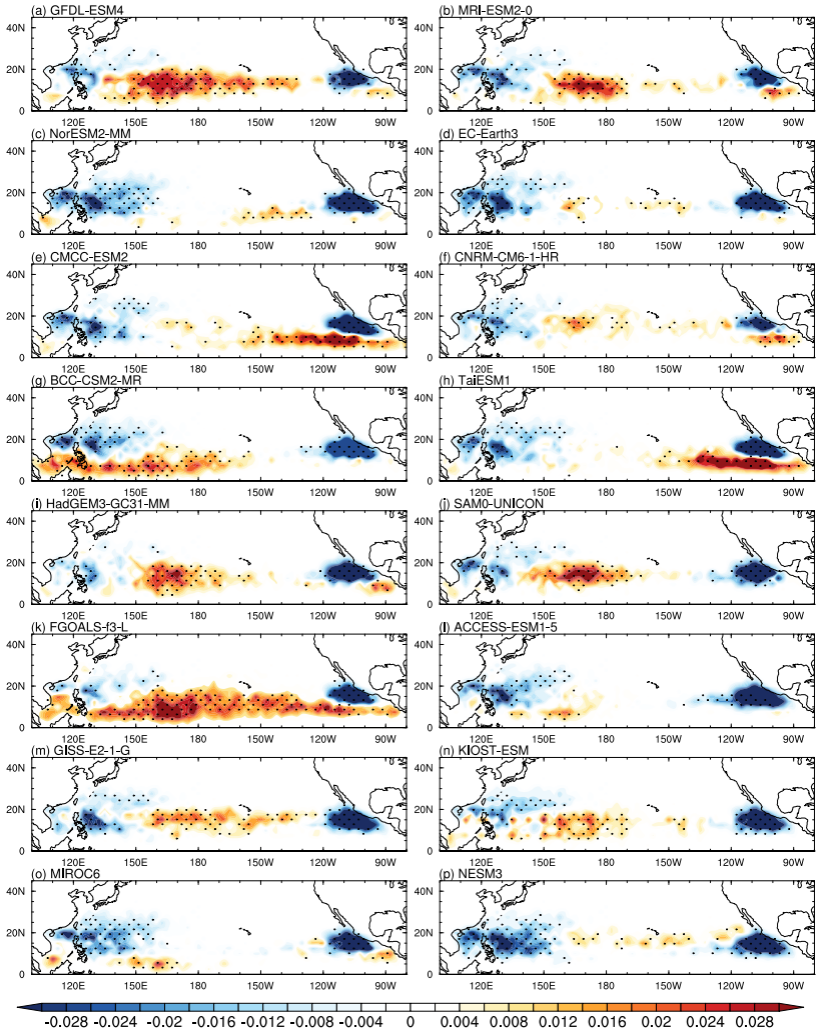

Fig. 2. Spatial distributions of climatological TCGD difference between 16 CMIP6 historical models and observation in JJASON during 1979–2014. Only values exceeding the 95% confidence level are stippled.

Figure 2 shows the JJASON TCGD difference between 16 historical models and observation over the North Pacific region during 1979–2014. The results show significant negative biases of TCGD in both WNP and ENP for most of the models, which is mainly due to coarse model resolution. Especially in the ENP, where a large number of TCs are concentrated in a very small basin in the observations, which is challenging to replicate in climate models with resolution lower than 100 km. Meanwhile, positive TCGD biases over CNP were also found in some historical models such as GFDL-ESM4, MRI-ESM2-0, HadGEM3-GC31-MM, SAM0-UNICON, FGOALS-f3-L, GISS-E2-1-G, KIOST-ESM, NESM3, BCC-CSM2-MR. In contrast, models like ACCESS-ESM1-5, MIROC6, TaiESM1, NorESM2-MM, CNRM-CM6-1-HR, EC-Earth3, CMCC-ESM2, exhibit TCGD closer to the

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

observation over the CNP, implying a smaller bias. It should be noted that in Fig. 1, we calculated the absolute TCGD, rather than the relative TCGD, so the positive TCGD biases over the CNP in some models that generated only a few TCs were not very large, but their positive relative TCGD biases over CNP still exist.

Since the resolutions of most of the 16 coupled models from historical experiments are coarser than 100km, it's difficult for these models to resolve TC structure or some other TC characteristics. Therefore, to examine whether this bias in the CNP could be reduced by improving the model resolution, we selected 10 high-resolution coupled models in hist-1950 experiments from HighResMIP and then calculated their TCGD.

b. TCGD in HighResMIP runs

Figure 3 presents the TCGD difference between 20 high resolution models from HighresMIP and observations. The left panel displays 10 coupled models from the hist-1950 experiment, while the right panel shows 10 corresponding atmospheric models from the highresSST-present experiment. However, we found that there are still a number of coupled models in hist-1950 runs that displayed positive TCGD biases over the CNP, especially HadGEM3-GC31-HM and CMCC-CM2-VHR4. It is worth noting that, when comparing different resolution models from the same institution (e.g., CNRM-CM6-1 and CNRM-CM6-1-HR), the increase in resolution does not reduce the positive TCGD biases of models in the CNP. Instead, it generates more TCs in the CNP owing to the overall increase in the simulated TC frequency due to the improved model resolution. As a result, negative biases in the ENP and WNP are reduced while the positive bias in the CNP is also amplified. Roberts et al. (2020) and Vecchi et al. (2019) suggested that the reason for overall increased TC frequency with higher resolution is a higher conversion rate of pre-TC "seeds" into TCs. Roberts et al. (2020) also pointed out that the large biases in CMCC and HadGEM3 families are due to their grid point dynamical schemes, while TC underestimation in EC-Earth3P and ECMWF families arises from their dynamical cores. These results further indicate that improving the model resolution only partially refine the TCGD bias over the CNP.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

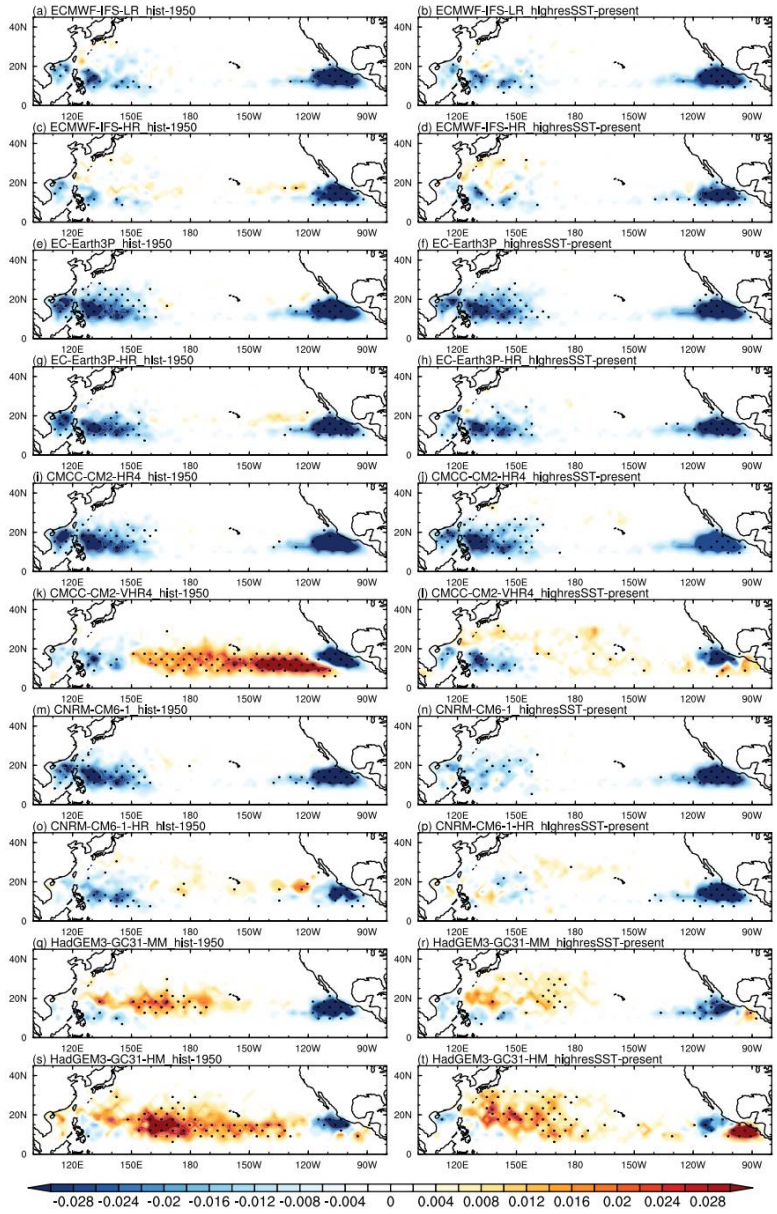

Fig. 3. Spatial distributions of climatological TCGD difference between 20 CMIP6-HighResMIP models and observation in JJASON during 1979–2014. Models in the left and right panels are from hist-1950 runs and highresSST-present runs, respectively. Only values exceeding the 95% confidence level are stippled.

To further explore the possible causes of the TCGD bias, we conducted a comparison of models in the highresSST-present experiment (right panel in Fig. 3), which were forced with observed SST. We found that for the same model, the positive TCGD bias in the CNP region was significantly reduced, although some models such as CMCC-CM2-VHR4 still displayed significant positive TCGD biases over the CNP. However, the biases in these models, as well as in other models with small TCGD biases over the CNP, were reduced obviously, while

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

TCGD increases in the WNP and ENP compared to hist-1950 simulations. As showed in Figs. 3 and 4, TC genesis regions were more concentrated in the WNP and ENP, closer to the observations. Such a phenomenon could be seen in almost all models, particularly those that had not previously accurately simulated the TCGD climatology of the CNP. An improvement in TC simulation in highresSST-present experiment may be due to the usage of observed SST forcing. This will be discussed in the following sections.

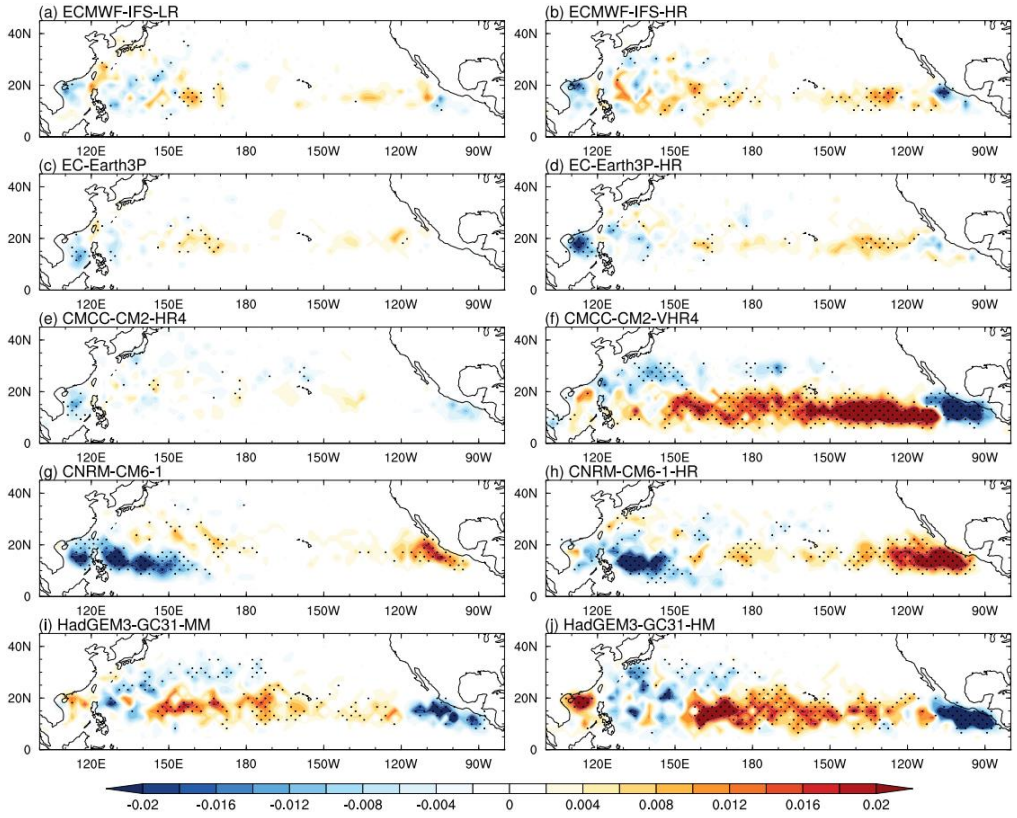

Fig. 4. Spatial distributions of climatological TCGD difference between coupled hist-1950 and uncoupled highresSST-present models in JJASON during 1979–2014. Only values exceeding the 95% confidence level are stippled.

Roberts et al. (2020) noted that the absolute frequency of TCs in models can be affected by different detection algorithms. Therefore, in our study, we compared the results extracted by TSTORMS with those of the simulated TC extracted by TRACK and TemptExtreme employed by Roberts et al. (2020), as illustrated in Fig. 5. From the Taylor diagrams of models' TCGD climatology over the North Pacific obtained using different TC detection algorithms. The root mean square error, standard deviation, and pattern correlation coefficients for each model are very close to 1, indicating similar spatial distribution of TCGD climatology obtained from the different detection algorithms.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

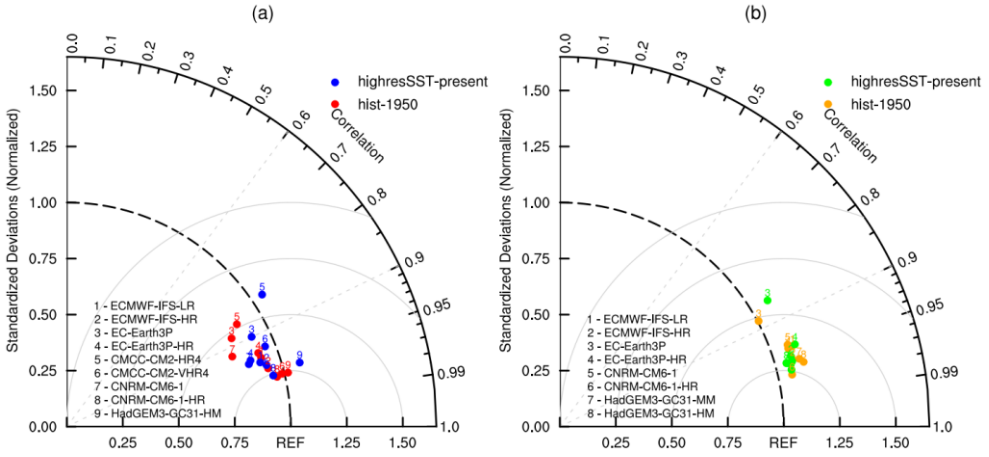

Fig. 5. Taylor diagram of monthly TCGD climatology over the North Pacific from hist-1950 and highresSST-present models using three different TC detection algorithms. The spatial correlation is plotted as the azimuthal angle, and the normalized variance is plotted as the radial distance from origin. The REF represents the TCGD calculated by TSTORMS, thus the distance between the point and REF represents the RMSE. The positions of the dots denote the magnitude of TCGD bias between TSTORMS and (a) TRACK, (b) TemptExtreme.

c. TCGD comparison between historical and AMIP simulations

In the comparison between high-resolution coupled and uncoupled atmospheric models discussed above, a significant improvement was observed in the simulation of TC genesis over the CNP for atmospheric models in Fig. 4. However, this improvement was not evident for some models with few absolute TC frequencies, only 4 out of 10 models displayed a clear shift in the sign and magnitude of the bias over the CNP. To make our findings more robust, we found 10 models corresponding to the coupled historical models in the uncoupled AMIP experiment, and analyzed whether this phenomenon was also present between the coarse-resolution coupled and atmospheric models (Table1). Similar to the high-resolution models, the same model from both historical and AMIP experiment uses the same detection thresholds to eliminate the influence of the detection algorithm and thresholds on the results.

As seen in Fig. 6, most of the low-resolution models exhibit a similar phenomenon to that of the high-resolution models in Fig. 4, with even a more obvious TCGD bias over the CNP. 8 out of 10 coupled models simulate more TCs in the CNP and western ENP compared to atmospheric models, while underestimating the frequency of TC genesis in the WNP and ENP. The distribution of TCGD simulated by the atmospheric models is closer to the observations compared to the coupled models. This further support our finding that the

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

coupling process of the model is more important in contributing to the TCGD biases over the CNP than an increase in model resolution. In the following sections, we will classify coupled and uncoupled models to investigate the underlying causes of this phenomenon.

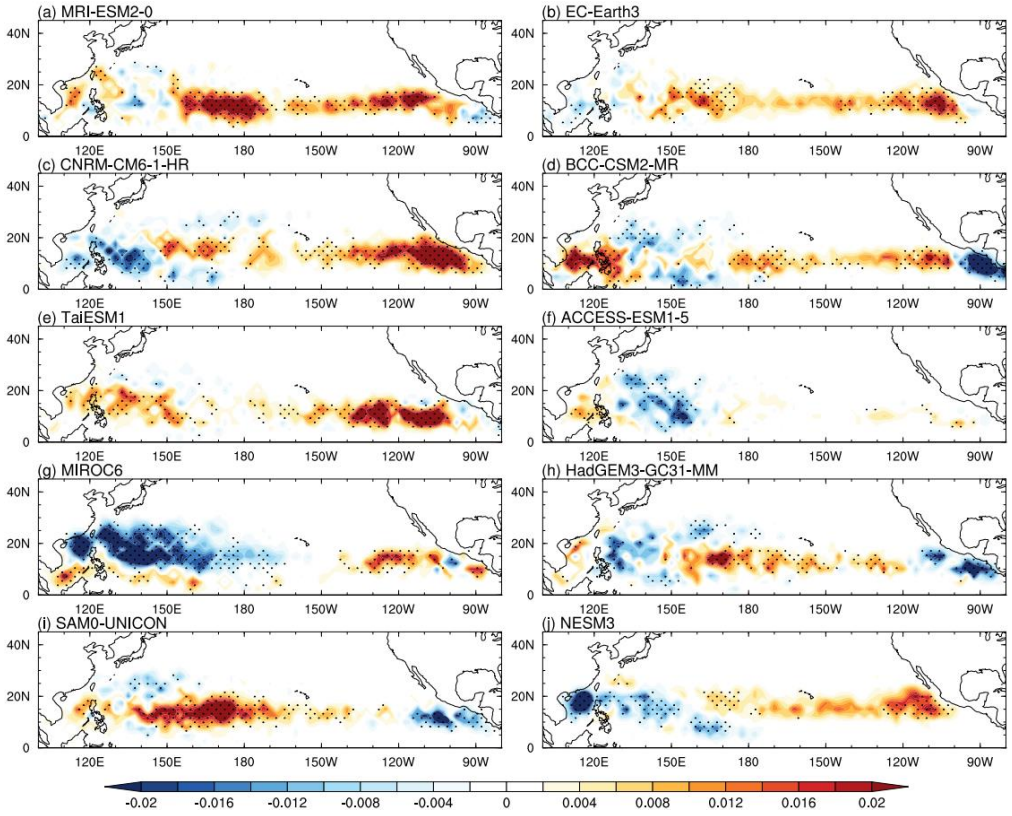

Fig. 6. Spatial distributions of climatological TCGD difference between 10 coupled historical and uncoupled AMIP models in JJASON during 1979–2014. Only values exceeding the 95% confidence level are stippled.

d. TCGD results of multi-model ensemble

For ease of comparison, we classified the models according to the magnitude of TCGD biases over the CNP in the model simulations. All the models from historical and hist-1950 experiments with positive TCGD biases commonly found over the CNP region were each divided into two groups. One group of models exhibits large TCGD biases in the CNP, while the other group shows smaller biases. The highresSST-present models with smaller biases were treated as a separate group, and resulting in a total of five groups labeled as G1–G5 (Table 3). Similar to the definition of relative TCGD, we calculated the TCGD of each model in the red box in Fig. 1 as a percentage of the TCGD of the entire North Pacific (Fig. 6), which is referred to as $TCGD_{CNP}$ ratio hereinafter. Consequently, the effect of the absolute

File generated with AMS Word template 2.0

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0224.1.

TC frequency of each model can be removed, and the magnitude of the bias can be quantified, reflecting the ability of each model to simulate the TC formation over the CNP. Based on this parameter, we then grouped the historical and hist-1950 models.

Experiments	Groups	Model name	$TCGD_{CNP}$ ratio (%)
observation		IBTrACS	3.39
historical	**G1**	GISS-E2-1-G	24.82
		FGOALS-f3-L	22.53
		GFDL-ESM4	20.56
		NESM3	19.57
		SAM0-UNICON	18.96
		MRI-ESM2-0	18.22
		HadGEM3-GC31-MM	17.63
		KIOST-ESM	17.34
		BCC-CSM2-MR	16.38
	G2	CMCC-ESM2	13.4
		EC-Earth3	13.28
		CNRM-CM6-1-HR	12.26
		NorESM2-MM	10.86
		ACCESS-ESM1-5	10.85
		TaiESM1	9.04
		MIROC6	5.26
hist-1950	**G3**	CMCC-CM2-VHR4	24.03
		HadGEM3-GC31-HM	21.48
		CMCC-CM2-HR4	17.49
		HadGEM3-GC31-MM	16.01
	G4	EC-Earth3P	13.77
		CNRM-CM6-1-HR	11.72
		CNRM-CM6-1	11.24
		EC-Earth3P-HR	9.67
		ECMWF-IFS-HR	6.6
		ECMWF-IFS-LR	5.85
highresSST-present	**G5**	CMCC-CM2-VHR4	15.09
		CMCC-CM2-HR4	12.76
		HadGEM3-GC31-HM	11.41
		HadGEM3-GC31-MM	9.15
		CNRM-CM6-1-HR	9.13
		CNRM-CM6-1	4.96
		ECMWF-IFS-LR	4.3
		EC-Earth3P-HR	3.89
		ECMWF-IFS-HR	3.7
		EC-Earth3P	2.29

Table 3. Classification of the CMIP6 models from three experiments used in this study and their $TCGD_{CNP}$ ratio values (%).

In Figure 7a, we present the $TCGD_{CNP}$ ratio values of 16 historical models and observation. We divided these models into two groups (G1 and G2) based on a 15% cut-off value of $TCGD_{CNP}$ ratio (indicated by the black dashed line), which was calculated using the method described above. The models with red bars represent those with large TCGD biases

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

over the CNP, while those with blue bars represent models with smaller biases. The $TCGD_{CNP}$ ratio value of the observation data is represented by the black bar and is the smallest among all historical models, indicating that even the model with the best simulation in $TCGD_{CNP}$ ratio in the historical experiment (~6%) still has a certain distance from the observation. The TCGD in these models still has a small positive bias over the CNP, while the $TCGD_{CNP}$ ratio is only about 3% in the observations. It is important to note that all the model has TCGD biases over the CNP, but the magnitude of the bias differs. The purpose of our study is to find out why such biases exist and how to reduce them, so it is necessary classify the model results into two groups. The classification criteria do not affect our conclusions but enable us to distinguish between models with large and small TCGD biases.

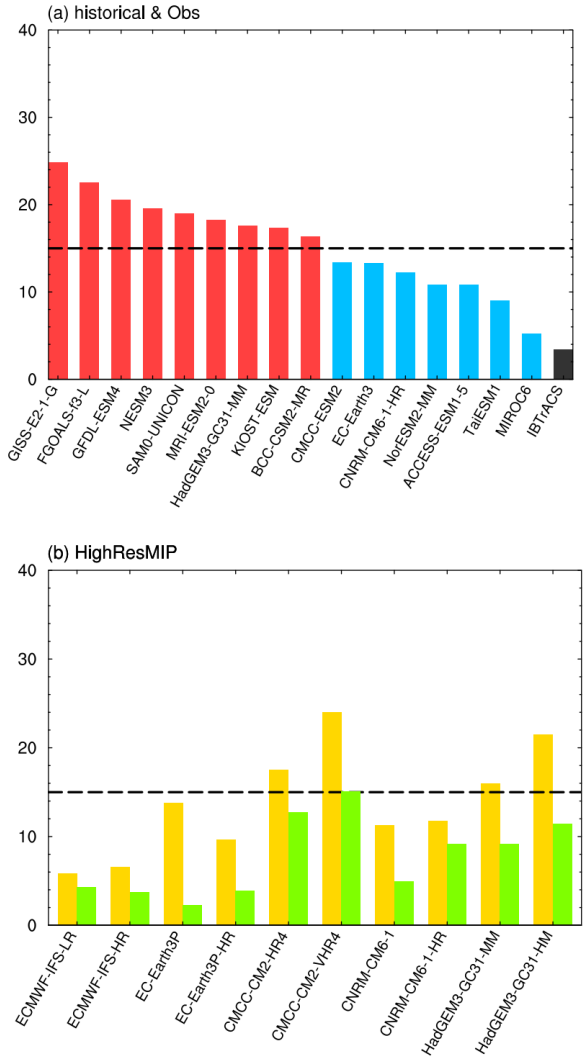

Fig. 7. $TCGD_{CNP}$ ratio (%) in (a) historical models and (b) HighResMIP models. The red bars represent models with large TCGD biases over the CNP grouped as G1, and the blue bars represent models with relatively small TCGD biases grouped as G2, the black bar represents the result from IBTrACS. The yellow bars represent models from coupled hist-

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

1950 runs, while the green bars represent highresSST-present runs. The black dash lines in (a) and (b) indicate the value of 15%.

Figure 7b shows the results of comparing the same model from two HighResMIP experiments. The results indicate that all highresSST-present models show a decrease in magnitude when compared with hist-1950 models, which indicates a significant improvement in TC simulations across these models. The $TCGD_{CNP}$ ratio in the CNP region is reduced, and the spatial distribution of TC is getting closer to the observation. Similarly, we applied the same classification method as the historical runs and divided the hist-1950 models into two groups. 4 out of 10 models with large TCGD biases over the CNP were grouped as G3, and remaining 6 small biased models were grouped as G4. And it is worth noting that, the $TCGD_{CNP}$ ratio of almost all the models in highresSST-present experiments (G5) are lower than 15%, only the CMCC-CM2-VHR4 is slightly larger than 15%, indicating a better performance. The specific classification and the magnitude of the $TCGD_{CNP}$ ratio can be found in Table 3.

By averaging models in these five groups, it can be seen from Fig. 8 that the models with large biases in TCGD over the CNP in the historical (G1) and hist-1950 (G3) experiments both exhibit positive biases over the CNP region. While there are almost no biases over the CNP in G2 and G4, despite the consistent underestimation of TC genesis frequency in the WNP and ENP in four groups. The multi-model ensemble mean of the highresSST-present simulation (G5) shows little TCGD bias in the CNP.

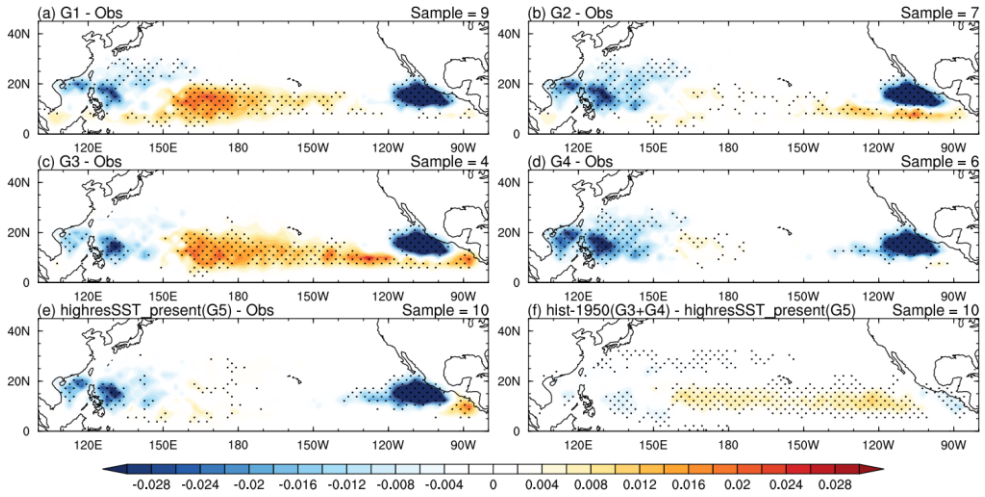

Fig. 8. Composite climatological JJASON TCGD difference between (a) Models with large TCGD biases over the CNP from historical experiment and observation. (b) As in (a), but for models with small TCGD biases. (c–d) As in (a–b), but for the hist-1950 experiment.

File generated with AMS Word template 2.0

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

(e) The difference between all 10 models from the highresSST-present experiment and the observation. (f) The difference between all 10 models from hist-1950 and highresSST-present. The number of composited models is indicated on the top-right corners of each plot. Only values exceeding the 95% significance level are stippled.

Because hist-1950 and highresSST-present simulations have the same model resolution, and physical parametrization, comparing the TCGD differences between the hist-1950 and highresSST-present simulations could largely exclude systematic errors induced by the above factors. Fig. 8f shows the composite difference of TCGD between the hist-1950 and highresSST-present. It was found that the negative TCGD biases over the WNP and ENP almost disappear, whereas the positive TCGD differences over the CNP still exist. This suggests that the negative TCGD biases over the WNP and ENP are mainly systematic errors, while the CNP TCGD biases may be related to the SST biases in the coupled simulations, since the two experiments differ in their SST forcing, which is related to air-sea interaction processes in coupled models. This will be further verified in the following sections.

Overall, it needs to be pointed out that most of the high-resolution hist-1950 models perform better in TC genesis simulation than the low-resolution historical models (Table 3). It further demonstrates that the increased resolution can indeed improve the TC simulation around the CNP (Figs. 1c and 1d). This can also be seen by comparing Figs. 2 and 3. But 4 out of 10 high-resolution models still have obvious TCGD biases (Fig. 3). Due to different absolute TC frequencies between different models, the $TCGD_{CNP}$ ratio gives a better indication of how each model simulates the CNP TC (Table 3).

The comparison between each hist-1950 and highresSST-present model in Fig. 7b and Table 3 further suggests that the source of such biases is not only related to resolution, but also to the differences between coupled and uncoupled models. The comparison between the high- and low-resolution models in hist-1950 experiment from the same institution (Fig. 4) also reveals that increasing the resolution actually amplifies the positive TCGD bias over the CNP. This indicates that the increase in resolution could not effectively reduce the TCGF bias. Comparing the coupled and uncoupled models demonstrates that the simulated SST bias may be the dominant reason for the TCGD bias.

4. Physical mechanisms

a. Large-scale environmental conditions

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

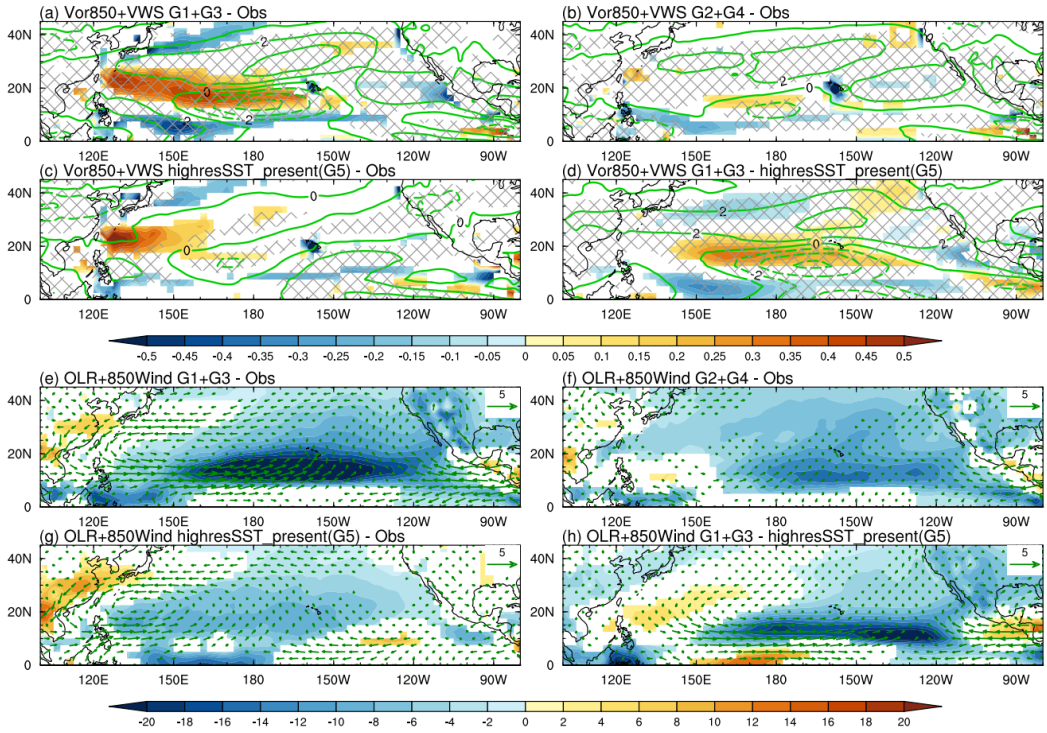

Fig. 9. Difference of (a) relative vorticity at 850 hPa (shading, unit: 10^{-5} s^{-1}) and vertical wind shear between 200 hPa and 850 hPa (contours, unit: m s^{-1}) between composite of 13 large-biases models from historical and hist-1950 experiment (G1 and G3) and the observation in JJASON from 1979 to 2014. (b) As in (a), but for 13 small-biases models (G2 and G4). (c) As in (a), but for all 10 models in the highresSST-present experiment (G5). (d) As in (a), but considering the difference between the composite of 13 large-biases models from historical and hist-1950 experiments and all 10 models from the highresSST-present experiment (i.e., (a–c)). (e–h) As in (a–d), but for outgoing longwave radiation (OLR, shading, unit: W m^{-2}) and 850-hPa winds (vectors, unit: m s^{-1}). Dashed contours represent the negative values, solid contours represent the positive values. Only values exceeding the 95% significance level are shown or hatched.

To further explore the cause of the TCGD biases over the CNP, large-scale environmental conditions related to TC genesis were further analyzed. While in the following part of the analysis of large-scale environmental conditions, for convenience, we combined models in G1 and G3, which exhibited large TCGD biases over the CNP, for comparison with observational data. Similarly, models in G2 and G4, with smaller biases, were also combined. Models in G5, which are atmospheric models, were analyzed as a separate category. Finally, we compared the results of the composites of G1 and G3 with G5.

The relative vorticity at 850 hPa (Vor850) and the vertical wind shear between 200 hPa and 850 hPa (VWS) were firstly analyzed (Figs. 9a-d). Both G1 and G3 show negative VWS anomalies over the CNP with positive Vor850 anomalies in the northwest compared with the

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

observations, which are favorable for TC genesis (Fig. 9a). In contrast, although the VWS and Vor850 anomalies in the G2, G4 and G5 are statistically significant, but they are much smaller than that in G1 and G3, indicating smaller biases in the large-scale environmental conditions (Figs. 9b,c).

In order to figure out the cause of the VWS and Vor850 anomalies, the 850-hPa horizontal wind was further analyzed. Figure 9e shows an anomalous cyclonic circulation from the WNP to the CNP in G1 and G3 at 850 hPa, which contributes to the positive Vor850 anomaly in Fig. 9a, along with a low-level westerly wind anomaly in the tropical WNP to CNP, while the wind in the tropical ENP is anomalously easterly. The wind field anomalies follow a Gill-type response induced by tropical diabatic heating (Gill, 1980). The cyclonic vorticity anomalies can be seen as a Rossby wave response, and the easterly wind anomaly was the Kelvin wave in Gill-type response. Overall, the low-level westerly wind anomaly reduces the background easterly wind, and the ascending motion triggers an anticyclonic divergence circulation at upper-level (Figs. 10a,e), causing an eastward displacement of the tropical upper-tropospheric trough (TUTT), which leads to strong vertical wind shear (Wang and Wu 2018). These dynamical environments could weaken the VWS in the CNP. However, the above environmental condition changes are not evident in G2, G4 and G5 (Figs. 9f,g), corresponding to significant but much smaller VWS and Vor850 anomalies (Figs. 9b,c). The difference between G1 (along with G3) and G5 shows an obvious cyclonic circulation in the tropical CNP (Fig. 9h), which is responsible for the positive TCGD biases in G1 and G3. This difference largely reduces the systematic errors.

Then, we further analyzed the outgoing longwave radiation (OLR; Figs. 9e-h) and vertical velocity at 500 hPa (Omega500; Figs. 10a-d). The negative OLR and Omega500 anomalies match well with the convergence zone of the low-level wind field in G1 and G3 (Figs. 9e and 10a). While the composite low-level wind field, OLR, and Omega500 in the CNP are small for G2, G4 and G5 (Figs. 9f,g and 10b,c), without showing a Gill-type response.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

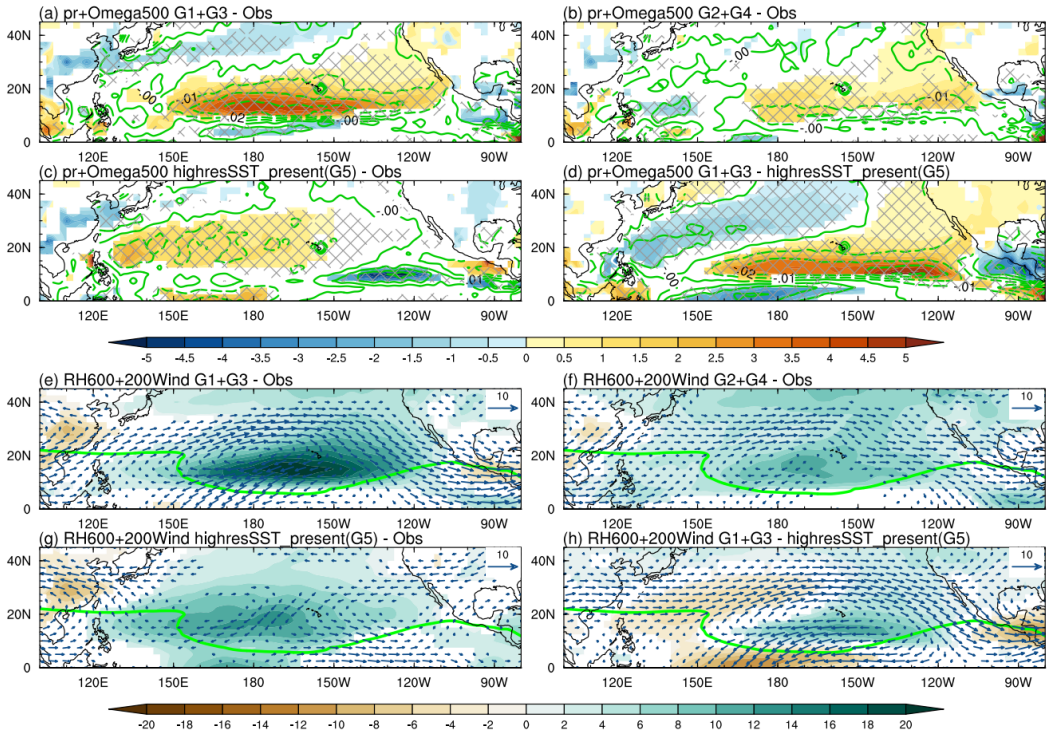

Fig. 10. (a–d) As in Figs. 9a–d, but for precipitation (shading, unit: 10^{-5} kg m^{-2} s^{-1}) and vertical velocity anomaly at 500hPa (contours, unit: hPa s^{-1}). (e–h) As in (a–d), but for relative humidity at 600 hPa (shading, unit: %) and 200 hPa-winds (vectors, unit: m s^{-1}), the green contours represent the climatological TUTT in observation. Only values exceeding the 95% significance level are shown or hatched.

To further seek the cause of the anomalous diabatic heating, SST simulations were evaluated. Recent studies have shown that relative SST (relative to the tropical mean SST) instead of absolute SST might be important to the TC genesis simulation in the models (Jonhson and Xie, 2010; Zhao et al., 2010; Murakami et al., 2011a). Here, Figure 11 shows the relative SST difference between the CMIP6 simulations and the observation. There is a positive SST bias extending from the northeastern Pacific to the tropical CNP in G1 and G3 (Figs. 11a,c), which may be related to the footprinting mechanism (Vimont et al. 2003). The warm SST bias over the ENP leads to the convergence of low-level wind field and weakened climatological trade winds, which causes precipitation and wind biases (Figs. 9e and 10a). This physical process is similar to that of the SSTA arising from the footprinting mechanism in Vimont et al. 2003. In contrast, G2 and G4 only show slight SST biases over the CNP (Figs. 11b,d). These results confirm that the positive SST biases in the CNP (red box in Fig. 11; 160°E–140°W, 5°N–15°N) for G1 and G3 may excite the Gill-type response, leading to a series of circulation anomalies in the CNP and creating favorable large-scale environmental

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0224.1.

conditions for TC formation. In contrast, the warm SST bias over the CNP largely reduces in G2 and G4 (Figs. 10b,c). In addition, the SST biases are accompanied by positive precipitation anomalies over the CNP, indicating intensification and displacement of the intertropical convergence zone (Fig. 10d), which is also a favorable environmental condition for TC genesis.

Finally, the composite 600hPa relative humidity anomalies also show that the models in G1 and G3 with large TCGD biases exhibit more obvious and concentrated regions of positive relative humidity anomalies in the middle troposphere (Fig. 10e). Conversely, the small biased coupled models and atmospheric models display more dispersed and weaker positive anomalies (Figs. 10f,g). This is easily understood since the Gill-type response to the positive SST bias induced an ascending motion in which the lifting and condensation of water vapor from the ocean makes the atmosphere wetter. From the analysis of various large-scale environmental fields and the multi-model ensemble mean results of the SST biases, we can also find that the bias is not only concentrated in the CNP, but also tends to extend towards the ENP. This also explains the positive bias over the western part of ENP in the TCGD comparison between coupled and uncoupled atmospheric models (Figs. 4 and 6).

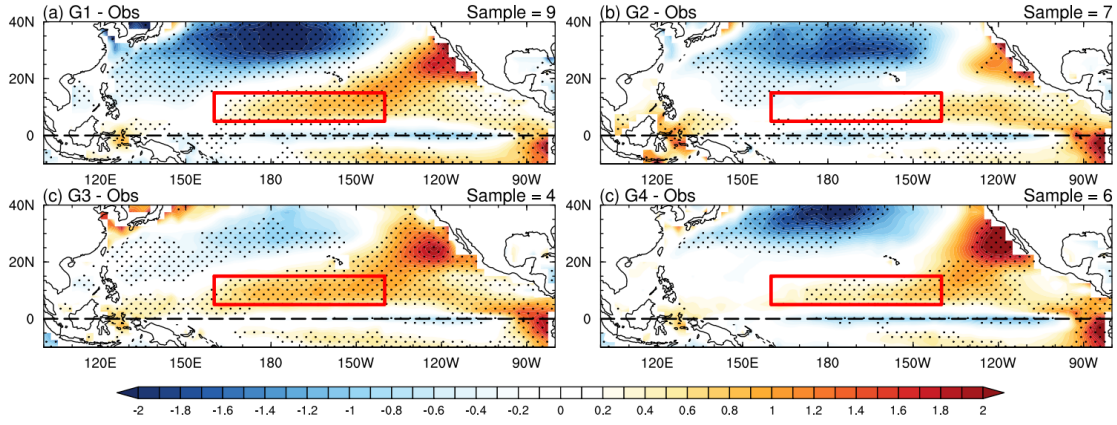

Fig. 11. Composite relative SST bias (shading, unit: K) between (a) large-biases models from the historical experiments and observation in JJASON from 1979 to 2014. (b) As in (a), but for small-biases models. (c–d) As in (a–b), but for models from the hist-1950 experiment. Only values exceeding the 95% confidence level are stippled based on the sign of difference.

b. Numerical experiments

To further validate the results, two numerical experiments were performed with HiRAM, which was forced by the prescribed SST. The control experiment was forced by the observed monthly SST from 1980 to 2009, similar to Zhao et al. (2020). The sensitive

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

experiment has the same SST forcing, but with the SST biases in the G1 and G3 simulations been exerted over 160°E–140°W and 5°N–15°N (red box in Fig. 11). Both experiments were integrated 30 years with 6-hourly outputs used to extract TCs and the monthly data to analyze the large-scale environmental conditions, respectively.

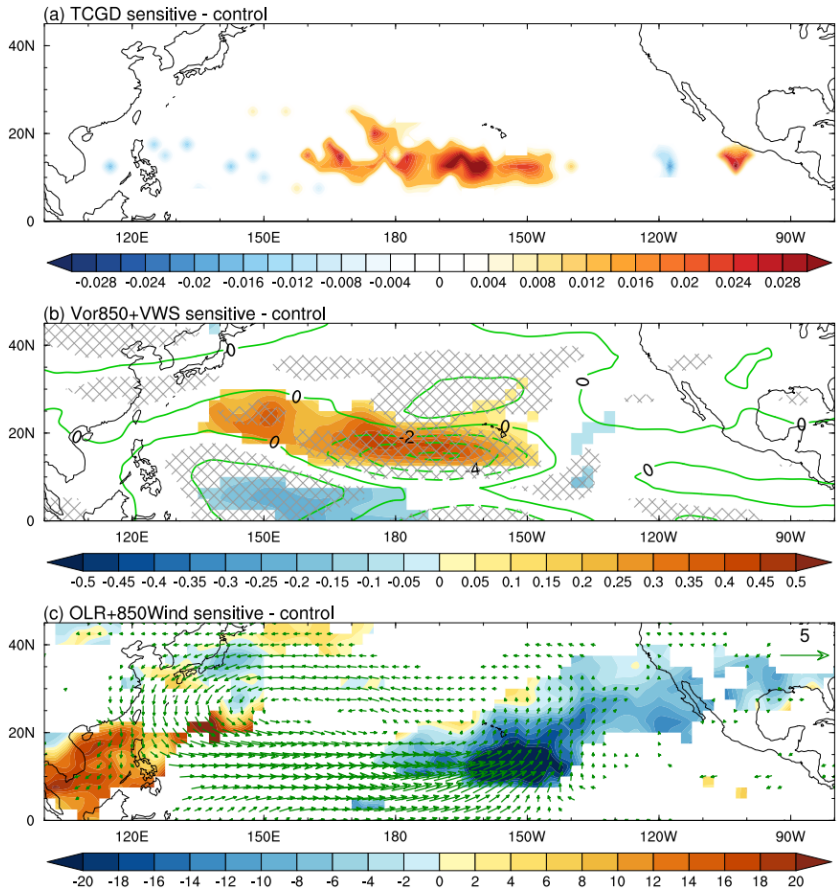

Fig. 12. Spatial distribution of the difference between the sensitive and control experiments in the (a) TCGD, (b) relative vorticity at 850 hPa (shading, unit: 10^{-5} s^{-1}) and vertical wind shear between 200 hPa and 850 hPa (contours, unit: m s^{-1}; dashed lines represent the negative values, solid lines represent the positive values), and (c) OLR (shading, unit: W m^{-2}) and 850-hPa winds (vectors, unit: m s^{-1}). Only values exceeding the 90% significance level are shown or hatched.

The TCGD difference between the sensitive and control experiments clearly shows a positive TCGD anomaly in the tropical CNP (Fig.12a), which is comparable to the CMIP6 results (Fig. 8f). As expected above, the large-scale circulation difference also shows a Gill-type response to the SST changes similar to those in the CMIP6 simulations (Figs. 12b and 9a; Figs. 12c and 9e). Nevertheless, the Kelvin wave response observed over the ENP is not statistically significant. There is a positive Vor850 anomaly along with a negative VWS anomaly originating from the anomalous heating, inducing a low-level cyclonic circulation

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0224.1.

and negative OLR in the tropical CNP (Figs. 12b,c). These findings are in accordance with the positive TCGD anomaly, further confirming the warm SST bias as the main cause of the positive TCGD bias in the CNP.

5. Conclusion and discussion

This study evaluated the simulation of TCs in the most recent CMIP6 climate models. Positive TCGD biases in the tropical CNP were found in some coupled models. We demonstrated that SST biases in the coupled models are the main cause of the TCGD biases. The physical mechanism underlying the relationship between SST biases and TCGD biases is depicted in Fig. 13. Specifically, a warm SST bias in the subtropical ENP extends to the tropical CNP, which induces a Gill-type response through increasing diabatic heating. This response triggers a Kelvin wave, leading to anomalous low-level easterly wind in the tropical ENP, and a cyclonic vorticity anomaly at the northwestern flank of the heating center through a Rossby wave response. The Rossby wave response also creates an anomalous low-level westerly flow in the southeastern WNP, which weakens the climatological easterly wind and extends the monsoon trough eastwards to the tropical CNP. Simultaneously, the ascending motion induced by SST warming generates anomalous anticyclonic divergence at the upper level, which brings about a wetter mid-troposphere, shifting the cyclonic TUTT accompanied by strong vertical wind shear to the east. The combined eastward shift of the TUTT and monsoon trough weakens the vertical wind shear over the tropical CNP, thus contributing to a favorable environment for TC genesis.

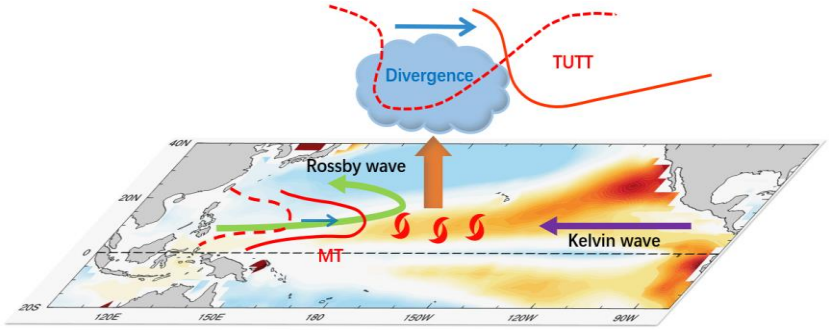

Fig. 13. Schematic illustration of the physical mechanisms for the overestimation of TC genesis over the CNP in CMIP6 coupled simulations. The shading indicates the multi-model ensemble mean SST biases in G1 and G3. The green arrow represents the Rossby wave response, and the purple arrow represents the Kelvin wave response. The red solid and dashed curves as well as the blue arrow over the WNP indicate an eastward shift of the monsoon trough. The red solid and dashed curves and the blue arrow at the top of the picture

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

represent the eastward shift of the tropical upper-tropospheric trough. The red TC-like symbols in the CNP indicate the positive TC genesis density biases. The thick orange arrow represents the upward motion, which corresponds to more rainfall (blue cloud) in this area.

The distribution of SST biases and the responses of the large-scale circulation in coupled models are pretty similar to those corresponding to the Pacific Meridional Mode (PMM) with interannual variability, which is described in Gao et al. (2018). Thus, we calculated the PMM index for each historical model following Chiang and Vimont (2004) to investigate whether it has a significant impact on TC genesis biases over the CNP in models. The correlation between the $TCGD_{CNP}$ ratio and the PMM index is not consistent throughout the models. In addition, it can also be seen from the figure that although the $TCGD_{CNP}$ ratio also has obvious interannual variations, systematic differences are also exists between the models. Indeed, some models, like HadGEM3-GC31-MM (r=0.58), show significant correlation between TCGD and PMM. There are still 12 out of the 16 models failed to reproduce such relationship (not shown). We hypothesize that their limited ability in simulating either PMM itself or the TCGD leads to the poor correlations. As we have put our main focus on the climatological biases of the CNP TCGD in this study, the interannual correlation between PMM and TCGD is beyond our scope. Thus, more studies on the TC-PMM relationship will be carried out in the future. On the other hand, although the SST bias and PMM may impact the large-scale environmental conditions over the CNP similarly, the TCGD bias caused by the SST bias in coupled models may not be directly linked to the PMM.

Recently, Gao et al. (2022) pointed out that in El Niño years, despite the warming of SST over the CNP, the TC frequency is not significantly different from La Niña years, which is due to the negative correlation between SST and 26℃ isothermal depths. But for the climatological mean, 14 out of the 16 models display a consistent temperature bias from the surface to subsurface ocean compared to observations (not shown). This suggests that the inconsistent surface and subsurface warming on interannual timescales has little impact on our conclusions. However, it is indeed worth further investigating the influence of surface and subsurface inconsistent warming biases on the interannual variability of the TC genesis in CMIP6 models.

The current results suggest that the SST simulation bias is one of the most important factors controlling the simulation of TC genesis over the North Pacific in climate models. As indicated in Zhang et al. (2023), the SST simulation changes from the cold bias to the warm

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

bias over the CNP in CMIP6, which is consistent with the warm SST biases of coupled models shown in our study. But it is worth pointing out that in CMIP5 models the CNP experienced a cold bias (Wang et al. 2014). Given the critical importance of the CMIP6 climate models in projecting TC activity, a clear understanding of model biases and their origin are quite crucial for accurately predicting, interpreting, and even reducing the uncertainty of future TC activity, especially under the background of global warming. Therefore, it is imperative to have a thorough understanding of the effects of SST biases on TC genesis simulations for accurate projections of future TC activity under different climate scenarios. Our results attach importance to further understanding the impact of SST biases on TC simulations, which is essential for improving future TC projections with these climate models.

Acknowledgments.

This study is jointly supported by the National Natural Science Foundation of China under Grants 42192555 and 42075003.

Data Availability Statement.

The tropical cyclone best track data used in this study is taken from the IBTrACS, version 4, which can be downloaded at https://www.ncei.noaa.gov/products/international-best-track-archive. The ERA5 reanalysis dataset can be downloaded at https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset. The HadlSST1 dataset can be downloaded at https://www.metoffice.gov.uk/hadobs/hadisst/. The CMIP6 data can be downloaded at https://esgf-node.llnl.gov/projects/cmip6/. The TC detection algorithm can be acquired at https://www.gfdl.noaa.gov/tstorms/. The HiRAM can be acquired at https://www.gfdl.noaa.gov/hiram-quickstart/.

REFERENCES

Balaguru, K., and Coauthors, 2020: Characterizing tropical cyclones in the energy exascale Earth System Model Version 1. *J. Adv. Model. Earth Syst.*, **12**, https://doi.org/10.1029/2019ms002024.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

Bell, S. S., S. S. Chand, S. J. Camargo, K. J. Tory, C. Turville, and H. Ye, 2019: Western north Pacific tropical cyclone tracks in CMIP5 models: statistical assessment using a model-independent detection and tracking scheme. *J. Climate*, **32**, 7191–7208, https://doi.org/10.1175/JCLI-D-18-0785.1.

Bengtsson, L., K. I. Hodges, and M. Esch, 2007: Tropical cyclones in a T159 resolution global climate model: comparison with observations and re-analyses. *Tellus*, **59A**, 396–416, https://doi.org/10.1111/j.1600-0870.2007.00236.x.

Burnett, A. C., A. Sheshadri, L. G. Silvers, and T. Robinson, 2021: Tropical cyclone frequency under varying SSTs in aquaplanet simulations. *Geophys. Res. Lett.*, **48**, e2020GL091980, https://doi.org/10.1029/2020GL091980.

Camargo, S. J., 2013: Global and regional aspects of tropical cyclone activity in the CMIP5 models. *J. Climate*, **26**, 9880–9902, https://doi.org/10.1175/jcli-d-12-00549.1.

Camargo, S. J., and A. A. Wing, 2015: Tropical cyclones in climate models. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 211–237, https://doi.org/10.1002/wcc.373.

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

Camargo, S. J., A. H. Sobel, A. D. Delgenio, J. A. Jonas, M. Kelley, Y. Lu, D. A. Shaevitz, and N. Henderson, 2016: Tropical cyclones in the GISS ModelE2. *Tellus*, **68A**, 31494, https://doi.org/10.3402/tellusa.v68.31494.

Camargo, S. J., and Coauthors, 2020: Characteristics of model tropical cyclone climatology and the large-scale environment. *J. Climate*, **33**, 4463–4487, https://doi.org/10.1175/JCLI-D-19-0500.1.

Chan, D., G. A. Vecchi, W. Yang, and P. Huybers, 2021: Improved simulation of 19th- and 20th-century North Atlantic hurricane frequency after correcting historical sea surface temperatures. *Sci. Adv.*, **7**, eabg6931, https://doi.org/10.1126/sciadv.abg6931.

Chand, S., K. J. Tory, H. Ye, and K. J. E. Walsh, 2017: Projected increase in El Niño-driven tropical cyclone frequency in the Pacific. *Nat. Climate Change*, **7**, 123–127, https://doi.org/10.1038/nclimate3181.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

Chand, S.S., Walsh, K.J.E., Camargo, S.J. et al., 2022: Declining tropical cyclone frequency under global warming. *Nat. Clim. Change*, **12**, 655–661 (2022). https://doi.org/10.1038/s41558-022-01388-4

Chen, K.-C., C.-H. Tsou, C.-C. Hong, H.-H. Hsu, and C.-Y. Tu, 2023: Effect of model resolution on simulation of tropical cyclone landfall in East Asia based on a comparison of 25- and 50-km HiRAMs. *Climate Dyn.*, https://doi.org/10.1007/s00382-023-06668-z, in press.

Chiang, J. C. H., and D. J. Vimont, 2004: Analogous Pacific and Atlantic Meridional Modes of Tropical Atmosphere–Ocean Variability. *J. Climate*, **17**, 4143–4158, https://doi.org/10.1175/JCLI4953.1.

Davis, C. A., 2018: Resolving tropical cyclone intensity in models. *Geophys. Res. Lett.*, **45**, 2082–2087, https://doi.org/10.1002/2017GL076966.

Doi, T., G. A. Vecchi, A. J. Rosati, and T. L. Delworth, 2013: Response to CO2 coubling of the Atlantic hurricane main development region in a high-resolution climate model. *J. Climate*, **26**, 4322–4334, https://doi.org/10.1175/jcli-d-12-00110.1.

Dutheil, C., and Coauthors, 2020: Impact of projected sea surface temperature biases on tropical cyclones projections in the South Pacific. *Sci. Rep.*, **10**, 4838, https://doi.org/10.1038/s41598-020-61570-6.

Gao, C., L. Zhou, C. Wang, I. I. Lin, and R. Murtugudde, 2022: Unexpected limitation of tropical cyclone genesis by subsurface tropical central-north Pacific during El Niño. *Nat. Commun.*, **13**, 7746, https://doi.org/10.1038/s41467-022-35530-9

Gao, S., L. Zhu, W. Zhang, and Z. Chen, 2018: Strong Modulation of the Pacific Meridional Mode on the Occurrence of Intense Tropical Cyclones over the Western North Pacific. *J. Climate*, **31**, 7739–7749, https://doi.org/10.1175/JCLI-D-17-0833.1.

Gill, A. E., 1980: Some simple solutions for heat-induced tropical circulation. *Quart. J. Roy. Meteor. Soc.*, **106**, 447–462, https://doi.org/10.1002/qj.49710644905.

Guo Y.-P., and Z.-M. Tan, 2018: Westward migration of tropical cyclone rapid-intensification over the Northwestern Pacific during short duration El Niño. *Nat. Commun.*, **9**, 1507, https://doi.org/10.1038/s41467-018-03945-y.

Guo, Y.-P., and Z.-M. Tan, 2022: Influence of track change on the inconsistent poleward migration of typhoon activity. *J. Geophys. Res. Atmospheres*, **127**, e2022JD036640, https://doi.org/10.1029/2022JD036640.

Haarsma, R. J., and Coauthors, 2016: High resolution model intercomparison project (HighResMIP v1.0) for CMIP6. *Geosci. Model Dev.*, **9**, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016.

Han, Y., M.-Z. Zhang, Z. Xu, and W. Guo, 2021: Assessing the performance of 33 CMIP6 models in simulating the large-scale environmental fields of tropical cyclones. *Climate Dyn.*, **58**, 1683–1698, https://doi.org/10.1007/s00382-021-05986-4.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis, *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hsu, W.-C., C. M. Patricola, and P. Chang, 2018: The impact of climate model sea surface temperature biases on tropical cyclone simulations. *Climate Dyn.*, **53**, 173–192, https://doi.org/10.1007/s00382-018-4577-5.

Johnson, N. C., and S.-P. Xie, 2010: Changes in the sea surface temperature threshold for tropical convection. *Nat. Geosci.*, **3**, 842-845. https://doi.org/10.1038/ngeo1008.

Kim, D., A. H. Sobel, A. D. Del Genio, Y. Chen, S. J. Camargo, M. Yao, M. Kelley, and L. Nazarenko, 2012: The tropical subseasonal variability simulated in the NASA GISS general circulation model. *J. Climate*, **25**, 4641–4659, https://doi.org/10.1175/JCLI-D-11-00447.1.

Kim, H., G. A. Vecchi, T. R. Knutson, W. G. Anderson, T. L. Delworth, A. Rosati, F. Zeng, and M. Zhao, 2014: Tropical cyclone simulation and response to CO2 doubling in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **27**, 8034–8054, https://doi.org/10.1175/JCLI-D-13-00475.1.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone best track data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

Knutson, T. R., J. J. Sirutis, M. Zhao, R. E. Tuleya, M. Bender, G. A. Vecchi, G. Villarini, and D. Chavas, 2015: Global projections of intense tropical cyclone activity for the late

twenty-first century from dynamical downscaling of CMIP5/RCP4.5 scenarios. *J. Climate*, **28**, 7203–7224, https://doi.org/10.1175/JCLI-D-15-0129.1.

Kossin, J. P., K. A. Emanuel, and S. J. Camargo, 2016: Past and projected changes in western North Pacific tropical cyclone exposure. *J. Climate*, **29**, 5725–5739, https://doi.org/10.1175/JCLI-D-16-0076.1.

LaRow, T. E., L. Stefanova, and C. Seitz, 2014: Dynamical simulations of North Atlantic tropical cyclone activity using observed low-frequency SST oscillation imposed on CMIP5 model RCP4.5 SST projections. *J. Climate*, **27**, 8055–8069, https://doi.org/10.1175/JCLI-D-13-00607.1.

Li, H., and R. L. Sriver, 2018: Tropical cyclone activity in the high- resolution Community Earth System Model and the impact of ocean coupling. *J. Adv. Model. Earth Syst.*, **10**, 165–186, https://doi.org/10.1002/2017ms001199.

Li, T., M. Kwon, M. Zhao, J.-S. Kug, J.-J. Luo, and W. Yu, 2010: Global warming shifts Pacific tropical cyclone location. *Geophys. Res. Lett.*, **37**, L21804, https://doi.org/10.1029/2010gl045124.

Liu, H., C. Wang, S.-K. Lee, and D. Enfield, 2012: Atlantic warm-pool variability in the IPCC AR4 CGCM simulations. *J. Climate*, **25**, 5612–5628, https://doi.org/10.1175/jcli-d-11-00376.1.

McDonald, R. E., D. G. Bleaken, D. R. Cresswell, V. D. Pope, and C. A. Senior, 2005: Tropical storms: representation and diagnosis in climate models and the impacts of climate change. *Climate Dyn.*, **25**, 19–36, https://doi.org/10.1007/s00382-004-0491-0.

Morim, J., M. Hemer, F. Andutta, T. Shimura, and N. Cartwright, 2020: Skill and uncertainty in surface wind fields from general circulation models: Intercomparison of bias between AGCM, AOGCM and ESM global simulations. *Int. J. Climatol.*, **40**, 2659–2673, https://doi.org/10.1002/joc.6357.

Murakami, H., R. Mizuta, and E. Shindo, 2011a: Future changes in tropical cyclone activity projected by multi-physics and multi-SST ensemble experiments using the 60-km-mesh MRI-AGCM. *Climate Dyn.*, **39**, 2569–2584, https://doi.org/10.1007/s00382-011-1223-x.

Murakami, H., B. Wang, and A. Kitoh, 2011b: Future change of western North Pacific typhoons: projections by a 20-km-Mesh Global Atmospheric Model. *J. Climate*, **24**, 1154–1169, https://doi.org/10.1175/2010jcli3723.1.

Murakami, H., and B. Wang, 2022: Patterns and frequency of projected future tropical cyclone genesis are governed by dynamic effects. *Commun. Earth Environ.*, **3**, 77, https://doi.org/10.1038/s43247-022-00410-z.

Nakamura, J., and Coauthors, 2017: Western North Pacific tropical cyclone model tracks in present and future climates. *J. Geophys. Res. Atmos.*, **122**, 9721–9744, https://doi.org/10.1002/2017JD027007.

Ogata, T., R. Mizuta, Y. Adachi, H. Murakami, and T. Ose, 2015: Effect of air-sea coupling on the frequency distribution of intense tropical cyclones over the northwestern Pacific. *Geophys. Res. Lett.*, **42**, https://doi.org/10.1002/2015gl066774.

Ogata, T., R. Mizuta, Y. Adachi, H. Murakami, and T. Ose, 2016: Atmosphere-Ocean coupling effect on intense tropical cyclone distribution and its future change with 60 km-AOGCM. *Sci. Rep.*, **6**, 29800, https://doi.org/10.1038/srep29800.

Patricola, C. M., D. J. Cassidy, and P. J. Klotzbach, 2022: Tropical oceanic influences on observed global tropical cyclone frequency. *Geophys. Res. Lett.*, **49**, https://doi.org/10.1029/2022gl099354.

Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, https://doi.org/10.1016/j.jcp.2007.07.022.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos.*, **108**, https://doi.org/10.1029/2002JD002670.

Richter, I., 2015: Climate model biases in the eastern tropical oceans: causes, impacts and ways forward. *Wiley Interdiscip. Rev.: Climate Change*, **6**, 345–358, https://doi.org/10.1002/wcc.338.

Roberts, M. J., and Coauthors, 2020: Impact of model resolution on tropical cyclone simulation using the HighResMIP–PRIMAVERA multimodel ensemble. *J. Climate*, **33**, 2557–2583, https://doi.org/10.1175/JCLI-D-19-0639.1.

Accepted for publication in *Journal of Climate*. DOI 10.1175/JCLI-D-23-0224.1.

Shaevitz, D. A., and Coauthors, 2014: Characteristics of tropical cyclones in high-resolution models in the present climate. *J. Adv. Model. Earth Syst.*, **6**, 1154–1172, https://doi.org/10.1002/2014ms000372.

Sharmila, S., K. J. E. Walsh, M. Thatcher, S. Wales, and S. Utembe, 2020: Real world and tropical cyclone world. Part I: high-resolution climate model verification. *J. Climate*, **33**, 1455–1472, https://doi.org/10.1175/JCLI-D-19-0078.1.

Song, K., J. Zhao, R. Zhan, L. Tao, and L. Chen, 2022: Confidence and uncertainty in simulating tropical cyclone long-term variability using the CMIP6-HighResMIP. *J. Climate*, **35**, 2829–2849, https://doi.org/10.1175/JCLI-D-21-0875.1.

Strazzo, S., J. B. Elsner, J. C. Trepanier, and K. A. Emanuel, 2013: Frequency, intensity, and sensitivity to sea surface temperature of North Atlantic tropical cyclones in best-track and simulated data. *J. Adv. Model. Earth Syst.*, **5**, 500–509, https://doi.org/10.1002/jame.20036.

Studholme, J., A. V. Fedorov, S. K. Gulev, K. Emanuel, and K. Hodges, 2022: Poleward expansion of tropical cyclone latitudes in warming climates. *Nat. Geosci.*, **15**, 14–28, https://doi.org/10.1038/s41561-021-00859-1.

Tory, K. J., S. S. Chand, J. L. McBride, H. Ye, and R. A. Dare, 2013: Projected changes in late-twenty-first-century tropical cyclone frequency in 13 coupled climate models from Phase 5 of the Coupled Model Intercomparison Project. *J. Climate*, **26**, 9946–9959, https://doi.org/10.1175/jcli-d-13-00010.1.

Tory, K. J., H. Ye, and G. Brunet, 2020: Tropical cyclone formation regions in CMIP5 models: a global performance assessment and projected changes. *Climate Dyn.*, **55**, 3213–3237, https://doi.org/10.1007/s00382-020-05440-x.

Vecchi, G. A., and Coauthors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27**, 7994–8016, https://doi.org/10.1175/jcli-d-14-00158.1.

Vecchi, G. A., and Coauthors, 2019: Tropical cyclone sensitivities to CO2 doubling: roles of atmospheric resolution, synoptic variability and background climate changes. *Climate Dyn.*, **53**, 5999–6033, https://doi.org/10.1007/s00382-019-04913-y.

Accepted for publication in *Journal of Climate.* DOI 10.1175/JCLI-D-23-0224.1.

Vimont, D. J., D. S. Battisti, and A. C. Hirst, 2003: The seasonal footprinting mechanism in the CSIRO general circulation models. *J. Climate*, **16**, 2653–2667, https://doi.org/10.1175/1520-0442(2003)016<2653:TSFMIT>2.0.CO;2.

Vitart, F., J. L. Anderson, J. Sirutis, and R. E. Tuleya, 2001: Sensitivity of tropical storms simulated by a general circulation model to changes in cumulus parametrization. *Quart. J. Roy. Meteor. Soc.*, **127**, 25–51, https://doi.org/10.1002/qj.49712757103.

Wang, C., and L. Wu, 2018: Projection of North Pacific tropical upper-tropospheric trough in CMIP5 models: implications for changes in tropical cyclone formation locations. *J. Climate*, **31**, 761–774, https://doi.org/10.1175/jcli-d-17-0292.1.

Wang, C., L. Zhang, S.-K. Lee, L. Wu, and C. R. Mechoso, 2014: A global perspective on CMIP5 climate model biases. *Nat. Climate Change*, **4**, 201–205, https://doi.org/10.1038/nclimate2118.

Wu, T., and Coauthors, 2021: BCC-CSM2-HR: a high-resolution version of the Beijing Climate Center Climate System Model. *Geosci. Model Dev.*, **14**, 2977–3006, https://doi.org/10.5194/gmd-14-2977-2021.

Zhang, G., H. Murakami, X. Yang, K. L. Findell, A. T. Wittenberg, and L. Jia, 2021: Dynamical seasonal predictions of tropical cyclone activity: roles of sea surface temperature errors and atmosphere–land initialization. *J. Climate*, **34**, 1743–1766, https://doi.org/10.1175/jcli-d-20-0215.1.

Zhang, Q., B. Liu, S. Li, and T. Zhou, 2023: Understanding models' global sea surface temperature bias in mean state: From CMIP5 to CMIP6. *Geophys. Res. Lett.*, **50**, e2022GL100888. https://doi.org/10.1029/2022GL100888.

Zhao, M., I. M. Held, and G. A. Vecchi, 2010: Retrospective forecasts of the hurricane season using a global atmospheric model assuming persistence of SST anomalies. *Mon. Wea. Rev.*, **138**, 3858–3868, https://doi.org/10.1175/2010mwr3366.1.

Zhao, M., I. M. Held, S.-J. Lin, and G. A. Vecchi, 2009: Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Climate*, **22**, 6653–6678, https://doi.org/10.1175/2009jcli3049.1.

Zhao, J., R. Zhan, Y. Wang, S.-P. Xie, and Q. Wu, 2020: Untangling impacts of global warming and Interdecadal Pacific Oscillation on long-term variability of North Pacific

tropical cyclone track density. *Sci. Adv.*, **6**, eaba6813, https://doi.org/10.1126/sciadv.aba6813.