Response to comments from Reviewer #1

We are grateful for Reviewer #1's constructive and insight comments. We have addressed all the comments in the revised manuscript. Our point-to-point responses are hereafter in Italic Font.

Major comments:

1. Logic of the manuscript: It would be better to collect the descriptions of observational datasets in a new section before "Results", instead of in each result subsection. The resolution of all the observational data used should also be marked.

Response: This point was a common concern of all three anonymous reviewers. It is also our major structural revision implemented in the revised manuscript. Actually, we added a new section "3.2 Data used for evaluations" to introduce all the observational datasets that we used our purpose of model assessment, together with their basic characteristics and properties.

Another structural revision is the inclusion of a new subsection "4.1 Global mean surface air temperature variations from 1950 to 2014" to assess the ability of our models in reproducing the historical evolution of global climate for the recent past. Figure 2 is new and all subsequent figures are shifted and re-numbered.

2. How did you compare the low- and high-resolution data on a lat-lon map? Do you interpolate from low to high or high to low? and Why? Similar to the observational data, all the methods used should also be introduced and summarized before showing the results.

Response: Generally speaking, when visual inspection is the purpose, we just plot observations and simulations on their native grid. But for quantitative assessment of model biases, we re-gridded simulations onto the corresponding observation's grid. In the revised manuscript, we added a paragraph of explanation at the beginning of Section 4: "Data analysis and visualization are generally on the original or native grid of observation and models. An exception is on the assessment of models' biases with contrast to observation. In this case, simulations are re-gridded onto the grid of corresponding observation" in line 492-495..

3. In Table 2 and related subsection, I think using the same period as CERES-EBAF product to evaluate the two-version models is better. What is the meaning of errors (how do you calculate it) in Table 2 and text?

Response: Indeed, it is more convincing to use the same period for model/observation inter-comparison. We thus adopted the common period of 2001-2014. Values in Table 2 denote the annual mean ±interannual standard deviation for 2001-2014.

4. I wonder why the period of 1971-2000 is used. According to the description of historical simulation of these two models (L356-359), they both ends at 2014 as recommended by CMIP6. So, using period of 1995-2014 should be better as more observational data are available.

Response: Following your suggestion, all the analyses are now changed to the reference period of 1995-2014.

5. Figure 7 uses a different color set to represent high and mid resolution models from Figure 2. For reading more easily, I recommend to make the color legend consistent throughout the manuscript.

Response: Colors are now consistent among relevant figure, i.e. black, red, and blue for observations, BCC-CSM2-HR, and BCC-CSM2-MR, respectively.

6. L463-464: Can you explain why HR model improves the DJF precipitation in the SPCZ? Is it controlled by resolution or parameterization? Such kinds of information are very helpful to other model developers.

Response: We think the cause is multiple but we tend to conclude that the physical parameterization is the primary cause. We added some explanations in the revised manuscript. "This systematic bias is evidently reduced in BCC-CSM2-HR, especially with weakened precipitation in the South Pacific Convergence Zone (SPCZ). The improvement of SPCZ precipitation in BCC-CSM2-HR might be attributed to the implementation of the UWMT scheme which improved the simulation of low-level clouds over the tropical eastern South Pacific (Lu et al., 2020b) and reduced warm biases there (Fig. 10c)" in line 591-596

7. Figure 8: Here I think you should use the period of 2001-2014.

Response: Modified. It is numbered to Figure 9.

8: Figure 9: Large biases in Kuroshio extension and North Atlantic in higher resolution model should be marked and give possible reasons. I wonder whether this bias is resulted from the coarse resolution of observation, viz. the observation is "wrong" here due to its low resolution.

Response: The observation is certainly not wrong, but it may be not enough, with its low spatial resolution, to reveal detailed structures for the Golf Stream and the Kuroshio current, including their extensions to eastern basins. We added some explanations "We also noted that a belt of warm SST biases in the Kuroshio extension and in the North Atlantic in both models (Figures 10b and 10c), especially in the high-resolution model. This bias may be partly resulted from the coarse resolution of HadISST data used, as SST near the Kuroshio shows strong temperature gradients with filamentous structures (Shi and Wang, 2020)" in line 627-631.

9: Figure 10: The color bar is weird. It is not easy to capture the relative magnitude, especially the areas with biases around zero value.

Response: Modified

10: Why does HR model improve the TC density in western Pacific but not in the North Atlantic? Any explanations?

Response: Many studies show that increasing atmospheric resolution is helpful to improve the simulation of TC. This is also the case in our simulations with a general improvement

of TC density in BCC-CSM2-HR. The tropical North Atlantic, however, does not show improvement in BCC-CSM2-HR. We think that this persistent bias is possibly caused by cold SST biases. We added some explanations in lines 691-700 in the revised text.

11: It is very interesting that the HR model can produce an excellent wind-pressure relation. Can you give a short physical explanation?

Response: Yes, we are also happy to see that our high-resolution model can well reproduce the relation between wind speed and surface pressure for detected tropical cyclones. We think that this should not be a fortuitous result, since many other studies (e.g. Murakami et al., 2012; Sugi et al., 2017; Vecchi et al., 2019) concluded on the importance of high resolution in simulating TC. Those studies also demonstrate that the maximum wind speed of TC simulated by a model with approximately 50 km resolution can reach up to 50~60 m s-1. We added some discussions in the revised manuscript in Section "4.4.1 Tropical Cyclones".

Here, we can further present a case of TC simulated in the western tropical Pacific in BCC-CSM2-HR, as shown in Figure S1. It corresponds to model calendar 2003-11-23 UTC18:00:00. There is a clear TC structure with circular sea level pressure isobars and strong winds around the TC eyewall. The maximum wind speed can reach 53.9 m/s with the minimum sea level pressure of 975.2 hPa.

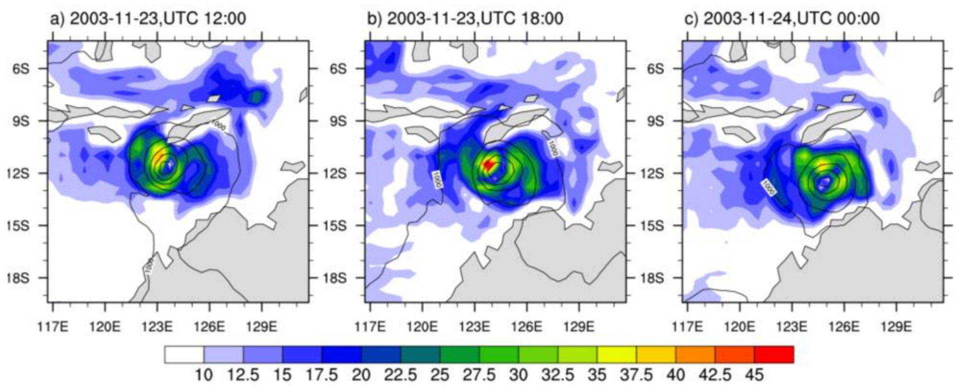

Figure S1. Snapshots of the TC with maximum intensities simulated by BCC-CSM2-HR in model date (a) 2003-11-23UTC12:00:00, (b) 2003-11-23UTC18:00:00 and (c) 2003-11-24UTC00:00:00, respectively. The shaded indicates the wind speed (m/s) at 10m. Contours indicate sea level pressure (hPa) with an interval of 5 hPa.

12: The color bar in Figure 17 should also be replaced by that or similar type used in Figure 16.

Response: The shaded areas in Figure 16 and 17 show OLR and zonal wind, respectively.

13: In Figure 18, the time series in (a)-(c) subpanels are not suitable for comparison. Maybe

you can use probability density function to show the asymmetry and skewness of ENSO.
Response: Following your suggestion, we added three maps to show skewness of ENSO.

14. How weaknesses of observational data could influence the model evaluation, especially for the high-resolution result is recommended to be discussed. For example, low resolution SST data is unable to capture the SST gradient along the Kuroshio and Gulfstream extension regions, it would be unfair for high resolution models if you use low resolution data as observational metrics.
Response: We absolutely agree with this point of view. When model's resolution reaches high levels, the lack of adequate observations becomes a major obstacle for our modeling efforts. We added some discussions in lines 627-631 in the revised manuscript.

Minor comments:

1. L3, P77: The following two papers are useful references here on how high resolution improves the monsoon simulation: Zhang L. et al. 2018. Effect of Horizontal Resolution on the Representation of the Global Monsoon Annual Cycle in AGCMs. Adv. Atmos.Sci., 10.1007/s00376-018-7273-9.https://link.springer.com/article/10.1007/s00376-018-7273-9 Yao J. et al. 2017: Improved performance of High-Resolution Atmospheric Models in simulating the East-Asian Summer Monsoon Rain belt. Journal of Climate, 30(21), 8825-8840, https://doi.org/10.1175/JCLI-D-16-0372.1 2.

Response: Modified

2. L51: Sea Surface Temperature (SST): the abbreviation should be used in Line 43 and the first letters should be in lower case.

Response: Modified.

3. P4, L102: In the climate model development community of China, the BCC holds a special position in that it is engaged in the development of its own climate models. The model has been used in both operational seasonal forecast and CMIP-like climate change simulation and projection. In contrast, other CMIP6 models from China are either hybrid models developed for research and education or purely research models. You may refer to Zhou et al. (2020) for the special position of BCC models in China: Zhou, T. et al. 2020: Development of Climate and Earth System Models in China: Past Achievements and New CMIP6 Results. J. Meteor. Res., 34(1), 1-19, doi: 10.1007/s13351-020-9164-0

Response: Modified.