**Response to comments from Reviewer #2**

***We thank reviewer #2 for his/her constructive and insight comments. We have addressed these comments in the revised manuscript. All our responses are in Italic Font.***

Major Comments:

1. Throughout the manuscript it is often unclear what grid(s) is being used for the analysis. That is, what grid are the observations and model results plotted on? Are they re-gridded to a common grid? Are they on the native grid? The interpretation of the results (or, in some cases, even the results themselves) can be influenced by this, I think it is important that the authors provide more detail on this topic.

***Response:***

***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.***

2. In Section 4 details of the observational datasets, including spatial resolution, temporal frequency, and time period, is not always clear. I would recommend that the authors reorganize the manuscript and include a Section that introduces all of the datasets and their details before Section 4.

***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.***

3. For all figures consistent line colors for observations, BCC-CSM2-MR and BCCCSM2-HR should be used. In particular, Figures 1, 2, 7, 8, 12, 14 and 18 use multiple different colors (red, blue, black, green, purple, etc.) to denote the simulations and observations. They should be the same colors for all figures and datasets.

***Response:***

***Modified. Line colors in all figures are replotted to consistence as black, blue, and red colors to denote observations, BCC-CSM2-MR, and BCC-CSM2-HR, respectively.***

4. Finally, I think the usefulness of this manuscript would be enhanced if the Conclusions (Section 5) included some discussion of how these main results compare to other work that has explored the effect of increases in resolution in climate models on the mean state, circulation, and phenomena of interest. In particular, the authors could put this work in the context of HighResMIP results, as well as some of the papers references in the Introduction of the manuscript.
***Response: Yes, this is a good point, but it is a little beyond the scope of our manuscript which modestly focuses on the documentation of the basic performance of our high-resolution model. But in the revised manuscript, we did add some discussions on JJA precipitation over the western Pacific warm pool, and on lack of TC activities over the North Atlantic, in comparison to other works. More simulations from other groups in the framework of HighResMIP start to be gradually released, which will allow us to do further diagnostics with multiple models.***

Specific Comments:
L29-31: Provide detail of the horizontal grid spacing here if possible.
***Response: Modified.***

L34: Consider changing "dynamic core" to "dynamical core" throughout the manuscript.
***Response: Modified.***

L57-58: Extend this sentence to put in context of the more recent CMIP6 (in addition to CMIP5).
***Response: Modified.***

L66: Remove "but."
***Response: Modified.***

L76: Change to "the QBO."
***Response: Modified***

L78: Replace "and" with ",".
***Response: Modified***

L94: Be specific that the authors are referring to atmosphere and ocean grids here.
***Response: Modified. That sentence is modified as "At present day, performing high-resolution climate simulations with model grid smaller than 50 km in the atmosphere and 0.25 °in the ocean is still a very costly effort …" in line 100-102.***

L98: HighResMIP is not "the primary activity" of CMIP6, as only a subset of models has completed it. Please reword.
***Response: Modified. That sentence is rewritten as "The High Resolution Model Intercomparison Project (HighResMIP, Haarsma et al., 2016) is a CMIP6-endorsed MIP***

***(Model Intercomparison Project), which aimed to investigate the impact of model resolution on climate simulation fidelity and systematic model biases." in line 105-108.***

Table 1: There seems to be a miss formatted Wu (2012) reference in the "Deep convection" row.

***Response: Modified.***

L371-L372: Three models? Only two models are introduced in Section 2.

***Response：Corrected.***

Figure 2: See major comment about regridding above.

***Response: We have added descriptions about how to plot for different resolution data.***

L403-408: Figure 3: It would be easier to see the biases if the models were plotted as a difference from the observations. Consider adding additional panels to the Figure.

***Response: Modified. We have added model biases with comparison against CERES-EBAF data.***

L446: Important variables in what way?

***Response: In the revised version, we have re-arrange this section. The expression of L446 in the first version of manuscript "Precipitation, land surface air temperature and sea surface temperature, sea-ice concentration are important variables of general concern" is now deleted to avoid ambiguity.***

Figure 6: See major comment about regridding above.

***Response：Data are plotted on their original resolution.***

L456-457: The authors should discuss the degradation in the simulation quality of precipitation east of the Philippines near the Pacific warm pool during JJA. I believe this was also seen in the Bacmeister et al. 2014 paper cited in the Introduction (see their Fig. 8). Some discussion of this degradation is needed here.

***Response:***

***We added some discussions about the precipitation degradation over the Pacific warm pool region as shown in Figure 6, now re-numbered as Figure 7 in the new manuscript. "In Figure 7f, we also noted that the amount of JJA precipitation in east of the Philippines and near the Pacific warm pool is worsened, since it is smaller in BCC-CSM2-HR than in BCC-CSM2-MR and GPCP data. This bias of lacking precipitation in BCC-CSM2-HR may partly be caused by a cold-SST bias over the western Pacific warm pool (Fig.10c)" in line 601-605.***

***We agree with the reviewer that this precipitation imperfection in high-resolution models is an important issue. With a simple bibliographic search, we found several other models reporting precipitation biases with resolution increased. So, we added more discussions in the section "5. Conclusions and discussions". Our work shows that enhancing resolution***

***does not noticeably improve climate mean state and a deterioration is even possible. For example, the decrease of JJA precipitation over the warm pool in our high-resolution model is still an important issue which certainly deserves further investigations with multiple models and simulations. Actually, other studies also reported similar issues. Haarsma et al. (2020) shows that increasing resolution in the EC-Earth model deteriorated the wet bias over the western Pacific warm pool. Bacmeister et al. (2014) analysed the high-resolution climate simulations performed with the Community Atmosphere Model (CAM), and showed that dry bias over the same region with enhanced resolution. Over the western Pacific warm pool, the atmospheric circulation and precipitation undergoes not only the impact of tropical variations such as MJO and TC, but also strong regional air-sea coupling.***

L465-468: This is an example where the underlying grid could be impacting the analysis (if the models and observations are not compared on a common grid – which is not obvious here).

***Response: Yes, we agree that different resolutions of observation and simulation affect the analyses. As mentioned in our response to Major comment #1, we added a clear statement on how to deal this issue of spatial resolution. Actually, when simulation is quantitatively compared against observation, we re-grid the simulation onto the grid of observation. But in Figure 7, the biases of precipitation over the tropical Pacific are real, not an artifact of grid transformation.***

Figure 8: See major comment about regridding above. The differences in grid could have implications here.

***Response: Two simulations were re-gridded to the grid of IMERG before processing..***

L491&L498-499: The authors could discuss how common this high-resolution cold bias is among other modeling groups, such as those that participated in HighResMIP.

***Response: Yes, we agree that this is another important issue for HighResMIP. Up to now, we cannot find publications to explore changes of warm SST bias in the eastern basins of subtropical oceans for those models participating in HighResMIP.***

Figure 10: The authors should include a panel of the observational (CRU) data at the top of this plot, similar to what was done for Figure 9.

***Response: We have added the CRU data plot in Figure 9 which is renumber to Figure 10 in revised version of the manuscript.***

L503-507: Is this somewhat to be expected since there are no wholesale (besides resolution) changes to the land modeling component?

***Response: Yes, the result conforms to expectation, since the land modeling component keeps very close to each other in the two models, and biases of near-surface air temperature over land are very similar to each other.***

L511-512: What is the resolution of the HadISST product? Please provide that information.
***Response: Provided.***

L544: What impact does this threshold have on the results if the same value is used for both resolutions?
***Response: In the first version of the manuscript, the threshold of relative vorticity at 850 hPa to detect TC for BCC-CSM2-MR and BCC-CSM2-HR were wrongly set to different values, which created difficulties for our purpose of inter-comparison, although that practice was perfectly valid to detect TC in individual models. We unified now the threshold (to a unique value of 15 ×10$^{-5}$ s$^{-1}$ in Figure 14). We also added some discussions on the influence of different thresholds: "If this threshold gets looser to 5×10$^{-5}$ s$^{-1}$, the averaged total global TC numbers per year in BCC-CSM2-MR and BCC-CSM2-HR will enhance to 55.9 and 101.1, respectively" in line 684-686. The following figure shows the global distribution, but omitted in the main text of the revised manuscript.***

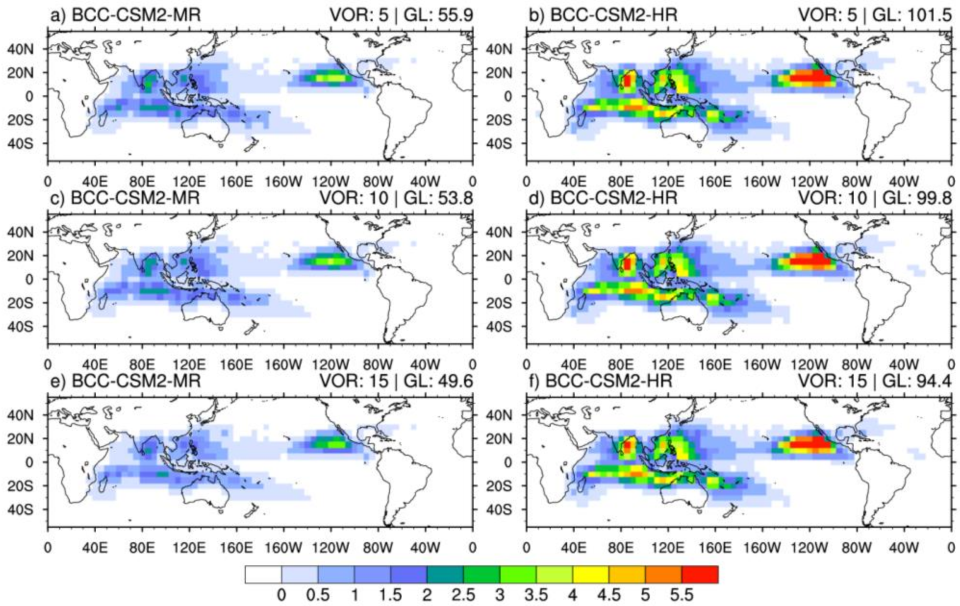


***Figure S1. The global distribution of tropical cyclone (TC) densities (number per year) averaged for BCC-CSM2-MR (left column), and BCC-CSM2-HR (right column). The thresholds of the relative vorticity for tracking the TC are 5 ×10$^{-5}$ s$^{-1}$, 10 ×10$^{-5}$ s$^{-1}$, 15 ×10$^{-5}$ s$^{-1}$, respectively. The value on the upper-right corner denotes the total number of global TCs on 5 °×5 °grid box.***

L542-551: Provide more information on the temporal frequency of the storm tracking? Is it daily for all tracking steps? Is intensity a mean or instantaneous?
***Response: Added more descriptions in "3.2 Data used for evaluations". The 1995–2014 6-hourly tropical cyclones observations from International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al., 2010) are used, which contains the information of latitude, longitude position, minimum central pressure, and maximum sustained winds at a***

***time frequency of every 6 hours of instantaneous values for all tropical cyclones. Following previous studies (Murakami, 2014), we use multiple criteria to detect TCs for every 6-hours interval outputs of instantaneous values from BCC-CSM2-HR and averaged means from BCC-CSM2-MR.***

Figure 14: This analysis is doing for daily storm intensities? Please provide more detail. How is the daily value (mean) calculated when storm intensities are typically represented instantaneously (as in IBTrACS). Are storms tracked daily or 6-hourly similar to IBTrACS? See comment above.

***Response: We added more description in "4.4.1 Tropical Cyclones" to explain for Figure 14 (now renumbered to Figure 15 in the revised version). "the maximum surface wind speed (minimum sea level pressure) of a given TC was defined as the instantaneous maximum (minimum) of the 6-hours interval in IBTrACS and BCC-CSM2-HR, but averaged value in BCC-CSM2-MR for wind speed at 10m (sea level pressure) …" in line 704-707.***

Figure 14: I find it difficult to believe that a model with 45 km grid spacing is replicating these high intensities (particularly, surface wind speeds) so well. But, it is also hard to interpret what a daily maximum intensity is. The authors should put this result into the context of the HighResMIP results, as well as Davis 2018 (https://doi.org/10.1002/2017GL076966).

***Response: We referenced to other similar simulations studies (e.g. Murakami et al., 2012; Sugi et al., 2017; Vecchi et al., 2019). 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. So, we have added some discussion in the "4.4.1 Tropical Cyclones". Here, we present a case of TC occurred on 2003-11-23, UTC18:00:00 in the western tropical Pacific in BCC-CSM2-HR as shown in Figure S2. There is a clear TC structure with the circular sea level pressure isobar and strong wind around the TC eyewall in the Figure S2, and the strongest TC in BCC-CSM2-HR, whose maximum wind speed can reach to 53.9 m/s with the minimum sea level pressure of 975.2 hPa (Figure S2).***

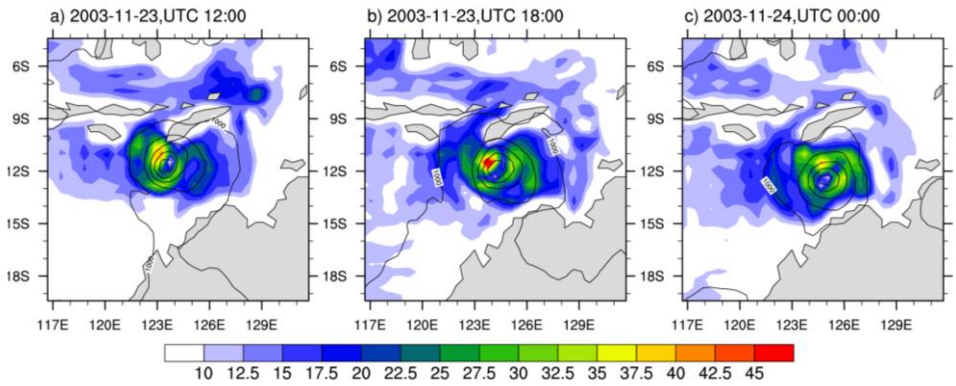

***Figure S2. Snapshots of the TC with maximum intensities simulated by BCC-CSM2-HR in model date (a) 2003-11-23UTC12:00:00, (b) 2003-11-23UTC18:00:00 and (c) 2003-11-24UTC00:00:00, respectively. The shaded indicates the wind speed (m/s) at 10m. Contours indicate sea level pressure (hPa) with an interval of 5 hPa.***

***The simulation of TC in tropical Atlantic is indeed a challenge for BCC-CSM2-HR. As your comment, we referenced to Davis (2018) which shows that models with horizontal grid spacing of one fourth degree or coarser should not produce a realistic number of category 4 and 5 storm in tropical Atlantic. So, we added some discussion about that in "5. Conclusions and discussions".***

L595: How is skillfully defined here?
***Response: rewritten that sentence. The signal of northward propagation is more evident in BCC-CSM2-HR than in BCC-CSM2-MR.***

L603: Provide detail of what observational dataset is used for OLR here.
***Response: Provided.***

L1310: Is it 3 hourly for BCC-CSM2-MR and BCC-CSM2-HR? To make this clear consider removing the "," after "2019."
***Response: Yes, in Figure 9, it is 3 hourly for BCC-CSM2-MR and BCC-CSM2-HR. we have rewritten the caption of Figure 9.***